

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:37:06 ; Search time 17 Seconds
(without alignments)
1363.013 Million cell updates/sec

Title: US-10-063-686-32

Perfect score: 445
Sequence: 1 MSGRPTIGLCTIALALSLA.....QDDKSLNKESTETQALVR 445

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size: 6

Total number of hits satisfying chosen parameters: 1355

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	1.8	147	1	LYC3 SHEEP
2	8	1.8	268	1	TRPA_HABIN
3	8	1.8	468	1	BMP3_RAT
4	8	1.8	472	1	BMP3_HUMAN
5	8	1.8	962	1	MLL9_MYCTU
6	7	1.6	89	1	YHUA_BACSU
7	7	1.6	113	1	PUR1_SYNP7
8	7	1.6	112	1	FRDC_PASMU
9	7	1.6	138	1	PA20_TRIFL
10	7	1.6	151	1	DKSA_BUCBP
11	7	1.6	161	1	LSA2_LYCES
12	7	1.6	165	1	RL19_DEIRA
13	7	1.6	185	1	FAPR_STRAM
14	7	1.6	186	1	FAPR_STRAE
15	7	1.6	207	1	Y214_MYCGE
16	7	1.6	236	1	AFPA_MOMO
17	7	1.6	247	1	COBS_SALTI
18	7	1.6	247	1	COBS_SALTY
19	7	1.6	294	1	HUPK_RHCCA
20	7	1.6	297	1	HTPX_SRRCA
21	7	1.6	298	1	Y025_MYCGE
22	7	1.6	302	1	GLT1_ECOLI
23	7	1.6	303	1	EST_Actin
24	7	1.6	337	1	GOGS_YEAST
25	7	1.6	341	1	BET3_CANAL
26	7	1.6	343	1	QUEA_BORBU
27	7	1.6	350	1	TME2_HUMAN
28	7	1.6	353	1	GBO_DROME
29	7	1.6	377	1	RIM2_YEAST
30	7	1.6	385	1	OP23_HABIN
31	7	1.6	386	1	OP24_HABIN
32	7	1.6	391	1	Y450_BUCAP

RESULT 1	ID	LYC3 SHEEP	STANDARD	PRT	147 AA.
AC	Q9TUN1	LYC3 SHEEP			
DT	28-FEB-2003	(Rel. 41, Created)			
DT	28-FEB-2003	(Rel. 41, Last sequence update)			
DE	28-FEB-2003	(Rel. 41, Last annotation update)			
OS	Lysozyme C 3 precursor (EC 3.2.1.17) (1,4-beta-N-acetylmuramidase C).				
OC	Ovis aries (Sheep).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Caprinae; Ovis.				
OX	NCBI_TaxID=9940;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=2008697; PubMed=10620405;				
RA	Wen Y, Irwin D.M.,				
RT	"Mosaic evolution of ruminant stomach lysozyme genes."				
RL	Mol. Phylogenet. Evol. 13:474-482(1999).				
CC	-1- FUNCTION: Lysozymes have primarily a bacteriolytic function; those in tissues and body fluids are associated with the monocyte-macrophage system and enhance the activity of immunogens.				
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan heteropolymers of the prokaryotes cell walls.				
CC	-1- SUBUNIT: Monomer.				
CC	-1- TISSUE SPECIFICITY: Expressed in stomach.				
CC	-1- MISCELLANEOUS: Lysozyme C is capable of both hydrolysis and transglycosylation; it shows also a slight esterase activity. It acts rapidly on both peptide-substituted and unsubstituted peptidoglycan, and slowly on chitin oligosaccharides.				
CC	-1- SIMILARITY: Belongs to family 22 of glycosyl hydrolases.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL, AF170559; AAD51637.1; -				
CC	EMBL, AF170556; AAD51637.1; JOINED.				
CC	EMBL, AF170557; AAD51637.1; JOINED.				
CC	EMBL, AF170558; AAD51637.1; JOINED.				
CC	HSSP; P00695; 1L26.				
DR	InterPro; IPR001916; Glyco_hydro_22.				
DR	Pfam; PF00062; 1ys; 1.				
DR	PRINTS; PRO0135; LY2LAC.				
DR	SMART; SMO0263; LY2L; 1.				
DR	PROSITE; PS00128; LACTALBUMIN LYSOZYME; 1.				
KM	Hydrolase; Glycosidase; Bacteriolytic enzyme; Digestion; Signal;				
KW	MultiGene family.				
FT	SIGNAL	1	18	BY SIMILARITY.	
FT	CHAIN	19	147	LYSOZYME C 3.	

ALIGNMENTS

34	415	1	ACRO_PIG	P08001	sus scrofa
35	420	1	PAG2_PIG	Q29079	sus scrofa
36	431	1	ACRO_RABIT	P48038	oryctolagus
37	435	1	LGNN_MOUSE	O89017	mus musculus
38	435	1	LGNN_RAT	Q97018	rattus norv
39	435	1	NHG2_PSEPU	O53552	pseudomona
40	448	1	Y347_CHLPN	O92616	chlamydia p
41	468	1	T10A_HUMAN	O00220	homo sapien
42	481	1	MORC_TREPA	O83361	treponema p
43	485	1	Y55K_BP22	P57019	bacterioph
44	489	1	FLIF_AQUAE	O67241	aquifex aeo
45	512	1	COX1_ARNSF	Q37705	artemia san

FT DISULFID 48 133 BY SIMILARITY.
 FT DISULFID 83 99 BY SIMILARITY.
 FT DISULFID 95 113 BY SIMILARITY.
 FT ACT SITE 53 53 BY SIMILARITY.
 FT ACT SITE 71 71 BY SIMILARITY.
 SQ SEQUENCE 147 AA; 16265 MM; 09601329F793EBD9 CRC64;

Query Match 1.8%; Score 8; DB 1; Length 147;
 Best Local Similarity 100.0%; Pred. No. 3.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 39 LVIGLLE 46
 Db 4 LVIGLLE 11

RESULT 2

TRPA_HABIN STANDARD; PRT; 268 AA.
 ID TRPA_HABIN STANDARD; PRT; 268 AA.
 AC P43753; 11
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Trypophan synthase alpha chain (EC 4.2.1.20).
 GN TRPA OR H1432.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellales; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KM20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kesteven A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McEweney K., Sutton G., Fitzhugh W., Fields C.A., Goode J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Corton M.D.,
 RA Ureackack T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fitchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
 RA Green C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd.";
 RL Science 269:496-512(1995).
 CC -1- FUNCTION: The alpha subunit is responsible for the aldol cleavage
 CC of indoleglycerol phosphate to indole and glyceraldehyde 3-
 CC phosphate.
 CC -1- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate
 CC = L-tryptophan + glyceraldehyde 3-phosphate.
 CC -1- PATHWAY: Tryptophan biosynthesis; fitch (lasc) step.
 CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By
 CC similarity).
 CC -1- SIMILARITY: Belongs to the trpa family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sb-sib.ch).

DR PROSITE, PS00167; TRP SYNTHASE ALPHA; 1.
 KW Tryptophan biosynthesis; lyase; Complete proteome.
 SQ SEQUENCE 268 AA; 28730 MM; E2FED0E50099E84 CRC64;

Query Match 1.8%; Score 8; DB 1; Length 268;
 Best Local Similarity 100.0%; Pred. No. 5.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 144 LSLGTAGA 151
 Db 225 LSLGTAGA 232

RESULT 3

BMP3_RAT STANDARD; PRT; 468 AA.
 ID BMP3_RAT STANDARD; PRT; 468 AA.
 AC P49002;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Bone morphogenetic protein 3 precursor (BMP-3).
 GN BMP3 OR BMP-3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Calvaria;
 RA Takao M., Hino J., Kangawa K., Matsuo H.;
 RL Submitted (Aug-1995) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 109-468 FROM N.A.
 RC TISSUE=Calvaria;
 RX MEDLINE=95186061; PubMed=7880444;
 RA Chen D., Feng J.Q., Feng M., Harris M.A., Mahy P., Mundy G.R.,
 RA Harris S.E.;
 RT "Sequence and expression of bone morphogenetic protein 3 mRNA in
 RT prolonged cultures of fetal rat calvarial osteoblasts and in rat
 RT prostate adenocarcinoma PA.iii cells."
 RL DNA Cell Biol. 14:235-239(1995).
 CC -1- FUNCTION: Induces cartilage and bone formation.
 CC -1- SUBUNIT: Homodimer; disulfide-linked (By similarity).
 CC -1- TISSUE SPECIFICITY: HIGH LEVELS IN TRACHEA, SPLEEN, SMALL
 CC INTESTINE, AND OVARY, AND AT LOW LEVELS IN CEREBELLUM, COSTA AND
 CC BONE MARROW.
 CC -1- SIMILARITY: Belongs to the TGF-beta family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sb-sib.ch).

EMBL; D63860; BAA09922.1; -;
 EMBL; S77492; AAB33725.1; -;
 DR HSSP; P12643; 3BMP.
 DR InterPro: IPR001839; TGPB.
 DR Pfam; PF00019; TGF-beta; 1.
 DR ProDom; PD000357; TGPB; 1.
 DR SMART; SM00204; TGF; 1.
 DR PROSITE; PS00250; TGF_BETA_1; 1.
 KW Signal; Growth factor; Cytokine; Osteogenesis; Chondrogenesis;
 KW Glycoprotein.
 FT SIGNAL 1 22 POTENTIAL.
 FT PROPEP 23 358 POTENTIAL.
 FT CHAIN 359 468 BONE MORPHOGENETIC PROTEIN 3.
 FT DISULFID 366 433 BY SIMILARITY.
 FT DISULFID 395 465 BY SIMILARITY.
 FT DISULFID 399 467 BY SIMILARITY.

FT CARBOHYD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 216 216 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 459 459 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 468 AA; 52675 MW; 05315D4954DC3CA1 CRC64;

Query Match
 Best Local Similarity 100.0%; Pred. No. 9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 SSLSILFF 224
 DB 439 SSLSILFF 446

RESULT 4

BMP3_HUMAN STANDARD; PRT; 472 AA.
 ID BMP3_HUMAN
 AC P12645;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Bone morphogenetic protein 3 precursor (BMP-3) (Osteogenin) (BMP-3A).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=89072730; PubMed=3201241;
 RA Wozney J.M., Rosen V., Celeste A.J., Mittleman B., Whitters M.J.,
 RA Kitz R.M., Hewick R.M., Wang E.A.;
 RT "Novel regulators of bone formation: molecular clones and
 RT activities";
 RL Science 242:1528-1534 (1988).
 CC -1- FUNCTION: Induces cartilage and bone formation.
 CC -1- SUBUNIT: Homodimer; disulfide-linked (by similarity).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN LUNG, OVARY AND SMALL INTESTINE.
 CC -1- SIMILARITY: Belongs to the TGF-beta family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M2291; AAA51836.1; -.
 DR PIR: D37218; BMR03.
 DR HSSP: P12643; 3BMP.
 DR GENBANK: HGNC:1070; BMP3.
 DR MIM: 112263; -.
 DR GO: GO:0005102; F:receptor binding; TAS.
 DR GO: GO:0007267; P:cell-cell signaling; TAS.
 DR GO: GO:0001501; P:skeletal development; TAS.
 DR InterPro: IPR001839; TGFb.
 DR Pfam: PF00019; TGF-beta; 1.
 DR ProDom: PD000357; TGFb; 1.
 DR SMART: SM00204; TGFb; 1.
 DR PROSITE: PS00250; TGF_BETA_1; 1.
 KW Signal; Growth factor; Cytokine; Osteogenesis; Chondrogenesis;
 KW Glycoprotein.

FT SIGNAL 1 22 POTENTIAL.
 FT PROPEP 23 362 POTENTIAL.
 FT CHAIN 363 472 BONE MORPHOGENETIC PROTEIN 3.
 FT DISULFID 370 437 BY SIMILARITY.
 FT DISULFID 399 469 BY SIMILARITY.
 FT DISULFID 403 471 BY SIMILARITY.
 FT DISULFID 435 436 INTERCHAIN (BY SIMILARITY).
 FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 472 AA; 53406 MW; 95C3B7ED5C9D596F CRC64;

Query Match
 Best Local Similarity 100.0%; Pred. No. 9.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 SSLSILFF 224
 DB 443 SSLSILFF 450

RESULT 5

MML9_MYCTU STANDARD; PRT; 962 AA.
 ID MML9_MYCTU
 AC P95235;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Putative membrane protein mmp19.
 DE Mmp19 OR RV2339 OR MT2402 OR MTCY98.08.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=9829587; PubMed=9634230;
 RA Cole S.T., Broeck R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglsmeier K., Gass S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellern S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206484; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft M., Hickey E.,
 RA Delcher A., Urdack T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL J. Bacteriol. 184:5479-5490 (2002).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: Belongs to the mmp1 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL: Z83860; CAB6156.1; -.
 DR EMBL: AB007080; AAK4696.1; -.
 DR PIR: D70661; D70661.
 DR TIGR: MT2402; -.
 DR TubercuList: RV2339; -.
 DR InterPro: IPR004707; ActII.
 DR InterPro: IPR004869; Mmp1.

```

DR Pfam, PF03176; MPM, 2.
DR TIGRfam; TIGR00833; act11, 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 25 45
FT TRANSMEM 213 233
FT TRANSMEM 256 276
FT TRANSMEM 302 322
FT TRANSMEM 335 355
FT TRANSMEM 383 403
FT TRANSMEM 768 788
FT TRANSMEM 796 816
FT TRANSMEM 820 840
FT TRANSMEM 867 887
FT TRANSMEM 896 916
FT TRANSMEM 937 957
SQ SEQUENCE 962 AA; 104612 MW; 9C8E3B4F44AC955 CRC64;

Query Match
Best Local Similarity 1.8%; Score 8; DB 1; Length 962;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 142 VLLSLGTA 149
DB 802 VLLSLGTA 809

RESULT 6
YHJA_BACSU STANDARD; PRT; 89 AA.
ID YHJA_BACSU
AC 00755;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein yhjA precursor.
GN yhjA OR BSU10440.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98240224; PubMed=9579061;
RA Azavedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brushchi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Comerford I.F., Cummings N.J., Daniel N.M.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Erttington J., Fabret C., Ferrati E., Foulger D.,
RA Filiz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gilm S.Y., Glaser P., Goffeau A., Gollightly B.J., Grand G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasaahara Y., Klastri-Blanchard W., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Maeda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,

```

```

RA Sato T., Scanlan E., Schleich S., Schroeder R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Togomi A.,
RA Tosato V., Uchiyama S., Vandenberg M., Vanlier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; Y14081; CAAT4463.1; -
DR EMBL; Z98109; CAB12884.1; -
DR PIR; A69833; A69833.
DR Subtilisin; BG13068; yhjA.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 27
FT CHAIN 28 89
FT SEQUENCE 89 AA; 9801 MW; 3980CF8D13EED66F CRC64;
SQ SEQUENCE 89 AA; 9801 MW; 3980CF8D13EED66F CRC64;

Query Match
Best Local Similarity 1.6%; Score 7; DB 1; Length 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 AVLLSLG 147
DB 6 AVLLSLG 12

RESULT 7
PURL_SYN7 STANDARD; PRT; 113 AA.
ID PURL_SYN7
AC Q55041;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphoribosylformyl-glycinimidase synthase II (EC 6.3.5.3) (FGAM
DE synthase II) (Fragment).
GN PURL.
OS Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
CX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RC
RX
RA Liu Y., Tsinoemas N.F., Golden S.S., Kondo T., Johnson C.H.;
RT Circadian expression of genes involved in the de novo purine
RT biosynthetic pathway of cyanobacteria."
RT Submitted (Aug-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CARBOLIC ACIDIVITY: ATP + N(2)-formyl-N(1)-(5-phospho-D-
CC ribosyl)glycinamide + L-glutamine + H(2)O = ADP + phosphate + 2-
CC (formimid) -N(1)-(5-phospho-D-ribosyl)acetamide + L-glutamate.
CC -1- PATHWAY: De novo purine biosynthesis; fourth step.
CC -1- SUBUNIT: Heterodimer of two subunits, purp and purl.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the FGAMS family.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```


Query Match 1.6%; Score 7; DB 1; Length 138;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 86 ITAVLV 92
 DB 6 ITAVLV 12

RESULT 10

DKSA_BUCBP STANDARD; PRT; 151 AA.

AC Q89AR3; 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Dnak suppressor protein homolog.
 GN DKSA OR BBP184.
 OS Buchnera aphidicola (subsp. *Baijzongia pistaciace*).
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Buchnera.
 CC NCBI_TaxID=135842;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426901; PubMed=12522265;
 RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
 RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,
 RA Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.,
 RT "Reductive genome evolution in *Buchnera aphidicola*."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).
 CC -1- SIMILARITY: Contains 1 dksa/lcrar-type zinc finger.
 CC -1- SIMILARITY: IN THE N-TERMINAL HALF, TO THE N-TERMINAL OF
 CC B. SUBTILIS YTEA.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AE014016; AAC26916.1; -
 DR InterPro; IPR000962; Znf_DsKa/TRAR.
 DR Pfam; PF01258; Zf_dksa/TRAR; 1.
 DR PRINTS; PR00618; DKSANZNFINGER.
 DR PROSITE; PS01102; DKSATRAR_ZN_FINGER; 1.
 KW Zinc-finger; Complete Proteome.
 FT ZN FING 114 138 TRAR/DKSA-TYPE (BY SIMILARITY).
 SQ SEQUENCE 151 AA; 17876 MW; 77B443C75CD558FA CRC64;

Query Match 1.6%; Score 7; DB 1; Length 151;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 217 SSSLTIF 223
 DB 10 SSSLTIF 16

RESULT 11

LAS2_LYCES STANDARD; PRT; 161 AA.

AC P13447;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Another specific LAT52 protein precursor.
 GN LAT52.

OS Lycopersicon esculentum (Tomato).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;

CC Lamids; Solanales; Solanaceae; Solanum.

OX NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. VF36;
 RX MEDLINE=89364704; PubMed=2770694;
 RA Twell D., Wing R., Yamaguchi J., McCormick S.;
 RT "Isolation and expression of an anther-specific gene from tomato."
 RL Mol. Gen. Genet. 217:240-245 (1989).
 CC -1- FUNCTION: May play a role during germination or early tube growth.
 CC -1- TISSUE SPECIFICITY: Expressed in anthers and pollen.
 CC -1- SIMILARITY: Belongs to the Ole e I family.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X15855; CAA33854.1; -
 DR FIR; S04765; S04765.
 DR InterPro; IPR006041; Ole_e1_ext.
 DR InterPro; IPR006040; PollenOle_e_1.
 DR Pfam; PF01190; PollenOle_e_1; 1.
 DR PROSITE; PS00925; OLEBI; 1.
 KW Glycoprotein; Signal.
 FT SIGNAL 1 17
 FT CHAIN 18 161 POTENTIAL.
 FT CARBOHYD 61 61 ANOTHER SPECIFIC LAT52 PROTEIN.
 SQ SEQUENCE 161 AA; 17834 MW; 7BD5E629F69758F2 CRC64;

Query Match 1.6%; Score 7; DB 1; Length 161;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 10 LCTIALA 16
 DB 11 LCTIALA 17

RESULT 12
 ID RL19 DEIRA STANDARD; PRT; 166 AA.
 AC Q9RMW4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE 50S ribosomal protein L19.
 GN RPLS OR DR0755.
 OS Deinococcus radiodurans.
 CC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 CC Deinococcaceae; Deinococcus.

OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson M.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile M., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Uretchack T., Zaleski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1."
 RL Science 286:1571-1577(1999).
 CC -1- FUNCTION: This protein is located at the 30S-50S ribosomal subunit
 CC interface and may play a role in the structure and function of the
 CC aminoacyl-tRNA binding site (by similarity).
 CC -1- SIMILARITY: Belongs to the L19p family of ribosomal proteins.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001931; AAF10334.1; -.
CC PIR; G75478; G75478.
CC PDB; INKW; 11-FEB-03.
CC TIGR; DR0755; -.
CC HAMAP; MF_00402; -.
CC InterPro; IPR001857; Ribosomal_L19.
CC Pfam; PF01245; Ribosomal_L19; 1.
CC PRINTS; PR00061; RIBOSOMAL19.
CC ProDom; PD002979; Ribosomal_L19; 1.
CC TIGRFAMs; TIGR01024; rplS_bact; 1.
CC DR PROSITE; PS01015; RIBOSOMAL_L19; FALSE_NEG.
CC K01 Ribosomal protein; Complete proteome; 3D-structure.
CC K01 SEQUENCE 166 AA; 18316 MW; 1DA8E740601FFAF9 CRC64;

```

```

Query Match 1.6%; Score 7; DB 1; Length 166;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 423 ARAQODK 429
DB 129 ARAQODK 135

```

```

RESULT 13
FAPR STRAP STANDARD; PRT; 185 AA.
ID FAPR STRAP
AC Q990P0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription factor fapR (Fatty acid and phospholipid biosynthesis
DE regulator).
GN FAPR OR SAV1228 OR SA1071 OR MW111.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699),
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
CX NCBI_TaxID=158878, 158879, 158820;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Karamori M., Matsunari H., Murayama A., Murakami H., Hoshiyama A.,
RA Mikutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus".
RU Lancet 357:1225-1240(2001).

```

```

RP SEQUENCE FROM N.A.
RC MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naito T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.,
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.".
RU Lancet 359:1819-1827(2002).
-1- FUNCTION: Transcriptional factor involved in regulation of
membrane lipid biosynthesis by repressing genes involved in fatty

```

```

CC -----
CC -1- SIMILARITY: Belongs to the fapR family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP003361; BAB57390.1; -.
CC DR EMBL; AP003133; BAB42323.1; -.
CC DR EMBL; AP004826; BAB94976.1; -.
CC PIR; G89895; G89895.
CC HAMAP; MF_01814; -.
CC K01 Fatty acid biosynthesis; Transcription regulation; Repressor;
CC K01 DNA-binding; Complete proteome.
CC K01 SEQUENCE 185 AA; 21434 MW; F779DC3C4D94E0F CRC64;

```

```

Query Match 1.6%; Score 7; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 97 LRRKIKL 103
DB 50 LRRKIKL 56

```

```

RESULT 14
FAPR STRAP STANDARD; PRT; 186 AA.
ID FAPR STRAP
AC Q8CSV2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription factor fapR (Fatty acid and phospholipid biosynthesis
DE regulator).
GN FAPR OR SE0903.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
CX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qian Z.-Q., Zhao G.-P., Ou D., Danchin A., Wen Y.-H.,
RA Yuan Z.-H., Zhao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228)".
RU Mol. Microbiol. 49:1577-1593(2003).
-1- FUNCTION: Transcriptional factor involved in regulation of
membrane lipid biosynthesis by repressing genes involved in fatty
acid and phospholipid metabolism (By similarity).
-1- SIMILARITY: Belongs to the fapR family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE016746; AA004500.1; ALT_INIT.
CC DR HAMAP; MF_01814; -.
CC K01 Fatty acid biosynthesis; Transcription regulation; Repressor;
CC K01 DNA-binding; Complete proteome.
CC K01 SEQUENCE 186 AA; 21421 MW; B817002658A461AE CRC64;

```

```

Query Match 1.6%; Score 7; DB 1; Length 186;
Best Local Similarity 100.0%; Pred. No. 42;

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 97 LRRRIKL 103
|||
Db 50 LRRRIKL 56

RESULT 15
Y214_MYCGE STANDARD; PRT; 207 AA.
ID Y214_MYCGE
AC P47456;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG214.
GN MG214.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
CX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne U.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Fuhmann J.L.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomo U.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lueker T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:197-403(1995).
CC -1- SIMILARITY: TO B.SUBTILIS YPH.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC
DR EMBL; U39701; AAC71433.1; -.
DR F04223; F64223.
DR TIGR; MG214; -.
DR InterPro; IPR005234; Cons_hypoth281.
DR Pfam; PF04079; DUF387; 1.
DR TIGRFAMs; TIGR00281; TIGR00281; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 207 AA; 23343 MW; 24741D38C557B56F CRC64;

Query Match 1.6%; Score 7; DB 1; Length 207;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 43 GLIFVCG 49
|||
Db 27 GLIFVCG 33

Search completed: May 27, 2004, 16:41:00
Job time : 19 secs

SEQUENCE CHARACTERISTICS:
LENGTH: 1389 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-702-344-20

Query Match 3.2%; Score 52.8; DB 1; Length 1389;
Best Local Similarity 49.1%; Pred. No. 2.4e-06;
Matches 169; Conservative 0; Mismatches 172; Indels 3; Gaps 1;

QY 856 TCTTCTACATCAAGAACCGTTGTGAAGGTCATTTTAATCTCTGTGTGAGATTC 915
DB 914 TCAGGTACCAACAGGCTCCCTGGCCCTTGGCCCTCATCTGCGCATTTGCGAGATCA 973
QY 916 CGAGATCATGTGTCATGTACATGCAAAACGACGTGAAGAACAGCAGCATGGTGCATTGT 975
DB 974 TCCGTGTGATCTCCAGTACCTGATCAAGCGCTGAAGCTGACAGAA---CAAGTTTG 1030
QY 976 CCAGTACCTGTTCGATGCTGCTAATGCTGTGTGTGCTTGAACAATACCTGTCTC 1035
DB 1031 CCAATGCTCATGACCTGCTCAATGCTGCTGCTGCTGCTGAGAAATTCATCAAT 1090
QY 1036 ATCTCAACGAATGATATCTATCAACTGCTATTAATGAGACAGATTTCTGTACATGAG 1095
DB 1091 TCCCTTAATGAGATGCTCAATCATCATGATGATCCATCAGGCAACAAATTCGCACTTGG 1150
QY 1096 CAAAGATGATCAAAATCTGTCCAAAGACTCAAGTCACTTATCATCTATTAATCTGCT 1155
DB 1151 CAGGAATGCTTCTTCCCTGCTCATGAGAAACATCATCAGAGTGGCTGCTGATTAAG 1210
QY 1156 TTGGAGACTTCAATATTTTCTAGAAAGCTTGTGTGTGTGT 1199
DB 1211 TTACTGACTTCTCTCTCTGTTGGGCAACCTTGTATGTGTGT 1254

RESULT 4
US-08-232-463-14

; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SRO ID NO: 14;
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-fls
US-08-232-463-14

Query Match 3.0%; Score 49.8; DB 1; Length 7218;
Best Local Similarity 2.5%; Pred. No. 7.7e-05;
Matches 9; Conservative 213; Mismatches 145; Indels 0; Gaps 0;

QY 8 TTCTTCACTATACCCACAGTCCAAAGACACTGCTGTCCAGGCTACAGTTCC 67
DB 1082 YY 1141
QY 68 TCCAGCAAGTCAATTCCTTATTAACGATGTGCTCAACACACTGAGTCACTAC 127
DB 1142 YY 1201
QY 128 CTAATTTGATCTGTTTGAATAAGATGTGACACCTCCACGAAATCTAAGTGAAT 187
DB 1202 YY 1261
QY 188 CATGTGGGAAGAGATACAATCCTTGCCGCTGATCTCGATGACCTGCTTGGC 247
DB 1262 YY 1321
QY 248 CATGATTTTACCTTGATGATCAACACCTTGCTGATCATTTGATTTGATGTGT 307
DB 1322 YY 1381
QY 308 TATTTGGGATTTGTTGTGTCGCGGTGTTTATGCTGATATGATGATGATGAT 367
DB 1382 YY 1441
QY 368 CGACCTC 374
DB 1442 ATTCCTC 1448

RESULT 5

US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Ubert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match 2.4%; Score 39.8; DB 4; Length 832;
Best Local Similarity 9.2%; Pred. No. 0.027;
Matches 32; Conservative 164; Mismatches 151; Indels 0; Gaps 0;

QY	ATCCCATTTTCTTTGGGTCCTCGGGGGGCTGTCGTCGAGGCTGGAACTCAGGA	638
Db	AGCCTTCTCCATCTAGACTCCTCGGCATGGCCCGTAGGTATCACTAGAGAGCA	10059
QY	GCTGCCCAAGGTTATGAAAGCGGCCCAAGTGAATATTAAGCCCTTTGGGGCATTTGGTAC	639
Db	GACCTTCAGAGATATGGCCCCCGAGATGATCTGGCCAAAGACCTTCATGCAGATCTCTCC	999
QY	ATGTGTCGTACATTTTATTTGGCTCAATCTGGACTAGGATTAATATCTTTCGGCCAG	699
Db	AATGGGGGCTTCATTACATGATAACTCAGGCTTGGTTATCAAGAGCCGCTTGGCCAG	939
QY	CAAAAGACTATAGCTGGGGCAGTGTGTTACTTGTATTTTCAACAGAACTAAAATGATCCT	759
Db	CATCACTAGTGACCTTGTACTCATTTACACATTTTTCACACATGATAGTAGGATATTC	879
QY	CCGTGATATCCCATCTTTGGTCCTCCATTCCTCTCTTCTTACATCA	819
Db	CATTAATATGATCATATGTGCGTAGCACCCTTGCTCTCTTCTCATCA	819

Db 1059 AGCCCTTCCATCATGTGACTCTCGGGCATGGCCCGTAGTGCATCCACTATGAGAGCA 1000

Qy 639 GCTGCCAGTTATGGAAGGGGCCAATGGAATATAAGCCCTTCCGGGATTCGGTAC 698

Db 999 GACTCTCAGAGATGGCCCCCAGATATATCGGGCAAGACCTTCATGCAGATCTCTC 940

Qy 699 ATGTGCTGTACCAATTATATTGGCCTCATCTGGAATTCATTCCTTCGTCGAC 758

Db 939 AATGGGGGCTTCATTACATGATTAACCTCAGGCTTGTTATCAGAGCGCCTTGGCCAG 880

Qy 759 CAATATGACTATAGCGGGGAGCTGGTTACTGTATTTTCAACAGAGATAAATGATCCT 818

Db 879 CATCACTATGAGACCTTGTACTATTACAAATTGTTTCCACAATGATGAGAGATATC 820

Qy 819 CCTGATCATCCATCTCTTGTGCTCTCTCATCTCTCTTCTTCAACATCA 869

Db 819 CATAAATATGATCAATGTGGCGTAGACCTTGTCTCTCTCCATCCAA 769

RESULT 85
 US-08-796-101-17/c
 ; Sequence 17, Application us/08796101
 ; Patent No. 6183752
 ; GENERAL INFORMATION:
 ; APPLICANT: EISEIN, STEPHEN E.
 ; APPLICANT: FINKEL, TOREN
 ; APPLICANT: SPEIR, EDITH
 ; APPLICANT: ZHOU, YI FU
 ; APPLICANT: ZHU, JIANHUI
 ; APPLICANT: ERDILE, LOREN
 ; APPLICANT: PINCUS, STEVEN
 ; TITLE OF INVENTION: RESTENOSIS/ATHEROSCLEROSIS DIAGNOSIS
 ; TITLE OF INVENTION: PROPHYLAXIS AND THERAPY
 ; NUMBER OF SEQUENCES: 184
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CURTIS, MORRIS & SAFFORD, P. C.
 ; STREET: 530 FIFTH AVENUE
 ; CITY: NEW YORK
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentm Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/796,101
 ; FILING DATE: 05-FEB-1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: KOMLASKI, THOMAS J.
 ; REGISTRATION NUMBER: 32,147
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 840-3333
 ; TELEFAX: (212) 764-5574
 ; INFORMATION FOR SEQ ID NO: 17:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1386 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-796-101-17

```

1      RESULT 9
2      US-09-085-273-53/c
3      ; Sequence 53, Application US/09085273
4      ; Patent No. 6267965
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Faioletti, Enzo
7      ; APPLICANT: Pincus, Steven E.
8      ; APPLICANT: Cox, William I.
9      ; APPLICANT: Kauffman, Elizabeth K.
10     ; TITLE OF INVENTION: RECOMBINANT POXVIRUS -
11     ; TITLE OF INVENTION: COMPOSITIONS AND USES
12     ; NUMBER OF SEQUENCES: 176
13     ; CORRESPONDENCE ADDRESS:
14     ; ADDRESSER: Curtiss, Morris & Safford
15     ; STREET: 530 Fifth Avenue
16     ; CITY: New York
17     ; STATE: New York
18     ; COUNTRY: United States of America
19     ; ZIP: 10036
20     ; COMPUTER READABLE FORM:
21     ; MEDIUM TYPE: Floppy disk
22     ; COMPUTER: IBM PC compatible
23     ; OPERATING SYSTEM: PC-DOS/MS-DOS
24     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
25     ; CURRENT APPLICATION DATA:
26     ; APPLICATION NUMBER: US/09/085,273
27     ; FILING DATE:
28     ; CLASSIFICATION:
29     ; PRIOR APPLICATION DATA:
30     ; APPLICATION NUMBER: US 08/471,014
31     ; FILING DATE: 06-JUN-1995
32     ; ATTORNEY/AGENT INFORMATION:
33     ; NAME: Frommer, Beq., William S.
34     ; REGISTRATION NUMBER: 25,506
35     ; REFERENCE/DOCKET NUMBER: 454310-2720
36     ; TELECOMMUNICATION INFORMATION:
37     ; TELEPHONE: (212) 840-3333
38     ; TELEFAX: (212) 840-0712
39     ; INFORMATION FOR SEQ. ID NO.: 53:
40     ; SEQUENCE CHARACTERISTICS:
41     ; LENGTH: 1386 base pairs
42     ; TYPE: nucleic acid
43     ; STRANDEDNESS: single
44     ; TOPOLOGY: linear
45     ; MOLECULE TYPE: DNA (genomic)
46     ;
47     ; US-09-085-273-53

```

Query Match	2.3%	Score 38.2;	DB 3;	Length 1386;
Best Local Similarity	45.7%;	Pred. No. 0.13;		
Matches 133;	Conservative	0;	Mismatches 138;	Indels 0;
				Gaps 0;

Query Match	2.3%	Score 38.2;	DB 3;	Length 1386;
Best Local Similarity	45.7%;	Pred. No. 0.13;		
Matches 133; Conservative	0;	Mismatches 158;	Indels 0;	Gaps 0;

Db 1059 AGCCTTCTCTCATCTGACTCTCGGCGATGCGCCGTAAGTCATCCACATGAGAGACA 1000
Qy 639 GCTGCCAGATTATGAAAGCGGCCAATGATATTAAGCCCTTTCCGGCATTCGGTAC 698
Db 999 GACTCTCAGAGATCGGCCCCCAGATGATGAGGCAAGACCTTCATGAGATCTCTC 940
Qy 699 ATGTGCTGACCATTTAATTTGCGCTCATCTGGAATTCATCTTGGCTCCAG 758
Db 939 AATGGCGGCTTCATTAACCTGATTAACCTGATGATGAGAGCGCCCTTGGCCAG 880
Qy 759 CAATGACTATAGCTGGGCGAGTGTACTTGTATTTCAACAGAAATTAATGATCCT 818
Db 879 CATCAGACTAGAGACTTGTACTCATTAACATTTTCCACATGATAGAGATATC 820
Qy 819 CTTGATCATCCATCTTTCGCTCTCTCCATTCCTTCTTACCATCA 869
Db 819 CATTAATATGATCAATGATGCGTGAAGACCTTGTCTCTCTCATCCAA 769

RESULT 10

US-09-916-963-53/C

Sequence 53, Application US/09916963

Patent No. 6632438

GENERAL INFORMATION:

APPLICANT: Paolletti, Enzo

PINCUS, Steven E.

Cox, William I.

Kaufman, Elizabeth K.

TITLE OF INVENTION: RECOMBINANT POXIVIRUS - CYTOMEGALOVIRUS,

NUMBER OF SEQUENCES: 176

CORRESPONDENCE ADDRESSES:

ADDRESS: Curtis, Morris & Safford

STREET: 530 Fifth Avenue

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/916,963

FILING DATE: 26-Jul-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/471,014

FILING DATE: 06-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Frommer Esq., William S.

REGISTRATION NUMBER: 25,506

REFERENCE/DOCKET NUMBER: 454310-2720

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 840-3333

TELEFAX: (212) 840-0712

INFORMATION FOR SEQ ID NO: 53:

LENGTH: 1386 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 53:

US-09-916-963-53

Query Match 2.3%; Score 38.2; DB 4; Length 1386;
Best Local Similarity 45.7%; Pred. No. 0.13;
Matches 133; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

Qy 579 ATCTCATTTTCTTCTGCGTCTCTGCGTCTGCTGCTGAGCTCGAATCGACAGA 638

Db 1059 AGCCTTCTCTCATCTGACTCTCGGCGATGCGCCGTAAGTCATCCACATGAGAGACA 1000
Qy 639 GCTGCCAGATTATGAAAGCGGCCAATGATATTAAGCCCTTTCCGGCATTCGGTAC 698
Db 999 GACTCTCAGAGATCGGCCCCCAGATGATGAGGCAAGACCTTCATGAGATCTCTC 940
Qy 699 ATGTGCTGACCATTTAATTTGCGCTCATCTGGAATTCATCTTGGCTCCAG 758
Db 939 AATGGCGGCTTCATTAACCTGATTAACCTGATGATGAGAGCGCCCTTGGCCAG 880
Qy 759 CAATGACTATAGCTGGGCGAGTGTACTTGTATTTCAACAGAAATTAATGATCCT 818
Db 879 CATCAGACTAGAGACTTGTACTCATTAACATTTTCCACATGATAGAGATATC 820
Qy 819 CTTGATCATCCATCTTTCGCTCTCTCCATTCCTTCTTACCATCA 869
Db 819 CATTAATATGATCAATGATGCGTGAAGACCTTGTCTCTCTCATCCAA 769

RESULT 11

US-08-658-665-54

Sequence 54, Application US/08658665

Patent No. 5997878

GENERAL INFORMATION:

APPLICANT: Paolletti, Enzo

PINCUS, Steven E.

Cox, William I.

Kaufman, Elizabeth K.

TITLE OF INVENTION: Recombinant Poxivirus - Cytomegalovirus,

NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESSES:

ADDRESS: Curtis, Morris & Safford, P.C.

STREET: 530 Fifth Avenue

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/658,665

FILING DATE: 05-JUN-1996

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Frommer Esq., William S.

REGISTRATION NUMBER: 25,506

REFERENCE/DOCKET NUMBER: 454310-2720.1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 840-3333

TELEFAX: (212) 840-0712

INFORMATION FOR SEQ ID NO: 54:

LENGTH: 3409 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-658-665-54

Query Match 2.3%; Score 38.2; DB 2; Length 3409;
Best Local Similarity 45.7%; Pred. No. 0.25;
Matches 133; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

Qy 579 ATCTCATTTTCTTCTGCGTCTCTGCGTCTGCTGCTGAGCTCGAATCGACAGA 638

Db	818	GACCTCAGAGGATGGGCCCCAGAAATGACTGGGCAAAAGCCTTACATGCAGATCTCTC	877
Qy	699	ATGTGTCGTAACATTTAATTGGCCCTCATCTGGACTAGTAATTCATCTTGGCTGGCAG	758
Db	878	AATGCGGCGGCTTATTACACTATAACCTCAGGCGCTGGTATTCAGAGGCGCGTTGGCAG	937
Qy	759	CAGATGACTATAGCTGGGGCAGGGTTCCTTGTTATTTTCAACAGAAAGAAAATGATCCT	818
Db	938	CATCAACAAGTACTCTTGTAACATTAACAATTTTCCAACAATGATGAGAGATATC	997
Qy	819	CGTATCATCCATCCCTTCCTCGTCTCTTCATCTCTTCTTCAACATCA	869
Db	998	CATTAATATGTATCAATGTGGGTAGACCTTGTCTCTCTCTCATCCAA	1048

RESULT 12

```

US-08-796-101-18
Sequence 18, Application US/08796101
Patent No. 6183752
GENERAL INFORMATION:
APPLICANT: ERSTEIN, STEPHEN E.
APPLICANT: FINKEL, TOREN
APPLICANT: SPEIR, EDITH
APPLICANT: ZHOU, YI FU
APPLICANT: ZHU, JIANHUI
APPLICANT: ERDLE, LORENE
APPLICANT: PINCUS, STEVEN
TITLE OF INVENTION: RESTENOSIS/ATHEROSCLEROSIS DIAGNOSIS
TITLE OF INVENTION: PROPHYLAXIS AND THERAPY
NUMBER OF SEQUENCES: 184
CORRESPONDENCE ADDRESS:
ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.
STREET: 530 FIFTH AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/796,101
FILING DATE: 05-FEB-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: KOWALSKI, THOMAS J.
REGISTRATION NUMBER: 32,147
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 764-5574
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3409 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-796-101-18

```

Query Match	2.3%;	Score 38.2;	DB 3;	Length 3409;
Best Local Similarity	45.7%;	Pred. No. 0.25;		
Matches 133;	Conservative	0;	Mismatches 158;	Indels 0;
				Gaps 0;

QY	Db
579 ATCTCATTTTCTTCGGGTCCTCTCGGGAGGTGTGCTGCTGAGCTGGGAATCAGAGA 63.8	579 AGCTTCTCTCATCTGACTCTCTGGGCGATGGCCCGTGTGATCATCACTGAGAGAGA 81.7
758	
639 GCTGCCAGATATGAGAGGGCGCCCAAGTGGAAATTAAGCCCTTTGGGCAATCGGTAC 65.8	818 GACTTCAGAGGATCGGCCCCCAGATGATCTGGCAAGAAGCTTCATGCAATCTCTC

Qy	699	ATGTGGTGGACCACTTAAATGGCCCTCATCTGACATGTAATTCATCTGGGTGGCAG	758
Db	878	AATGGCGCGCTTCATTACCTGTAATACCTCAGCGCTGGTATATCAGAGCGCGCTTGGCGCAG	937
Qy	759	CAATGACATATAGCTGGGCGAGTGGTACTTGTATTATTCACAGAGAATAAATGATCCT	818
Db	938	CATCACACTAGGACCTTGTACTCACTTACATCAACATTTGTTCCACAGATGATGAGGATATC	997
Qy	819	CCGATGATCCCATCTTTCGTCCTCTCCAAATTCCTCTTCTTCAACATCAA	869
Db	998	CATTAATATGTATCAATGTGGGTGAGACCTTGCTCTCTCCCTCATCCAA	1048

RESULT 13

```

1      US-09-085-273-54
2      : Sequence 54, Application US/09085273
3      : Patent No. 6267965
4      :
5      : GENERAL INFORMATION:
6      :
7      : APPLICANT: Paolietti, Enzo
8      : APPLICANT: Pincus, Steven E.
9      : APPLICANT: Cox, William I.
10     : APPLICANT: Kauffmann, Elizabeth K.
11     : TITLE OF INVENTION: RECOMBINANT POXVIRUS - CYTOMEGALOVIRUS,
12     : TITLE OF INVENTION: COMPOSITIONS AND USES
13     : NUMBER OF SEQUENCES: 176
14     :
15     : CORRESPONDENCE ADDRESS:
16     : ADDRESSEE: Curtis, Morris & Safford
17     : STREET: 530 Fifth Avenue
18     : CITY: New York
19     : STATE: New York
20     : COUNTRY: United States of America
21     : ZIP: 10036
22     :
23     : COMPUTER READABLE FORM:
24     : MEDIUM TYPE: Floppy disk
25     : COMPUTER: IBM PC compatible
26     : OPERATING SYSTEM: PC-DOS/MS-DOS
27     : SOFTWARE: Patent Release #1.0, Version #1.30
28     :
29     : CURRENT APPLICATION DATA:
30     : APPLICATION NUMBER: US/09/085,273
31     : FILING DATE:
32     :
33     : CLASSIFICATION:
34     : PRIOR APPLICATION DATA:
35     : APPLICATION NUMBER: US 08/471,014
36     : FILING DATE: 06-JUN-1995
37     :
38     : ATTORNEY/AGENT INFORMATION:
39     : NAME: Frommer Bsg., William S.
40     : REGISTRATION NUMBER: 25,506
41     : REFERENCE/DOCKET NUMBER: 454310-2720
42     : TELECOMMUNICATION INFORMATION:
43     : TELEPHONE: (212) 840-3333
44     : TELEFAX: (212) 840-0712
45     :
46     : INFORMATION FOR SEQ ID NO: 54:
47     :
48     : SEQUENCE CHARACTERISTICS:
49     : LENGTH: 3409 base pairs
50     : TYPE: nucleic acid
51     : STRANDEDNESS: single
52     : TOPOLOGY: linear
53     :
54     : MOLECULE TYPE: DNA (genomic)
55     :
56     : US-09-085-273-54

```

Query Match	2.3%;	Score 38.2;	DB 3;	Length 3409;
Best Local Similarity	45.7%;	Pred. No. 0.25;		
Matches 133;	Conservative 0;	Mismatches 158;	Indels 0;	Gaps 0;

Qy	578	ATCCATATTTCTTTCTGGAGTCTCTGGGTGCTGTGCTGAGCCCTGGGAACTGCACGA	638
Db	758	AAGCTCTTCCCTCATGTGACTCTGGGGATGGCCCGTAAGTCACTCCACTAGSAGAGCA	817
Qy	639	GCTGCCCAAGTATAGTAAGGCGGCAGTAGTAATAAGCCCTTTGGGCATTGGGTAC	698
Db	818	GACTCTCAAGGATGGCCCCCAGAAATGTACTGGGCAAAAGACCTTCACTGAGATCTTCTTC	877
Qy	699	ATGTGATGTAACCAATTAATTTAGGCTCATCGACTAGTAATTCATCTCTGGTGCACG	758

Tue Jun 1 09:37:56 2004

us-10-063-686-31.rni

Page 8

Db 878 AATGGCGCGCTTCATTACACTGTAACTCAAGCTGGTTATCAGAGGCGCGTTGGCCAG 937

Qy 759 CAATGACTCTAGCTGGGCGAGTGGTTATTTTCAACAGAAAGTAAAGATCCT 818

Db 938 CATCAACACTGATACCTTGACTCATTAACAATTGTTCACACATGATGAGAGATTC 997

Qy 819 CCTGATCATCCCATCCTTTCGTCTCTCCATCTCTTCTTCAACATCAA 869

Db 998 CATTAATATGTGATCAATGAGGTGAGCACTTGTCCTCTCTCATCAA 1048

RESULT 14
US-09-916-963-54
; Sequence 54, Application US/09916963
; Patent No. 6632438
GENERAL INFORMATION:
; Application Date: 07/04/09

Query Match	2.3%	Score	38.2	DB	4	Length	3409
Best Local Similarity	45.7%	Pred.	No.0.25				
Matches	133	Conservative	0	Mismatches	158	Indels	0
						Gaps	0

D_b 878 AATGGCGCGCTTCATTCACATGATTAACCTCAGCGCTGGTTTTCAGAGGCCCGCTTGGCCAG 93

O_y 759 CAATGACCTATAGCTGGGCGAGGGTCTCTGTTATTTCAACAGAGTAAATATGATCCT 818

D_b 938 CATCACACATGATCCTTCTACTCATTCACATTTCTTTCCACATGATGAGGAGATATC 997

O_y 819 CCTGATCATCCATCCCTTTCGTCCTCTCCATCTCTTCTTCAACATCAA 869

D_b 998 CATATAATATGATCATGTGCGTGAGCACCTTGTCTCTCCATCCAA 1048

RESULT 15
US-10-204-708-49
; Sequence 49, Application US/10204708
; Patent No. 6677731

Query Match	2.3%	Score 38.2	DB 4	Length 6306
Best Local Similarity	48.8%	Pred. No. 0.4		
Matches 103; Conservative	0	Mismatches 108	Indels 0	Gaps 0

Search completed: May 29, 2004, 07:46:53
Job time : 150 secs

Tue Jun 1 09:37:59 2004

us-10-063-686-32.rsp

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:22:44 / Search time 18 Seconds
(without alignments)
1287.290 Million cell updates/sec

Title: US-10-063-686-32

Perfect score: 2315
Sequence: 1 MSGRDTLGLCTLALSLA.....QDDKHSURNEGTGLQAVR 445

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	126	5.4	372	1	CYB_TRYBO
2	119	5.1	523	1	NU4M_TRYBO
3	118	5.1	497	1	NU4M_TRYBO
4	112.5	4.9	348	1	GAIR_MOUSE
5	110	4.8	495	1	NU4M_MOUSE
6	109	4.7	501	1	TLCR_MOUSE
7	108.5	4.7	346	1	GAIR_MOUSE
8	108.5	4.7	608	1	VATT_TRYBO
9	108	4.7	474	1	Y294_MOUSE
10	106	4.6	497	1	LMP2_MOUSE
11	106	4.6	677	1	ST21_TRYBO
12	104.5	4.5	475	1	Y294_MOUSE
13	103	4.4	465	1	NP1_MOUSE
14	103	4.4	518	1	FLC1_MOUSE
15	102.5	4.4	306	1	CD4_MOUSE
16	102	4.4	378	1	CD6_MOUSE
17	102	4.4	422	1	RBN_MOUSE
18	102	4.4	433	1	NU4M_MOUSE
19	102	4.4	446	1	DADR_MOUSE
20	102	4.4	446	1	DADR_MOUSE
21	101.5	4.4	491	1	NU4M_MOUSE
22	101.5	4.4	495	1	NU4M_MOUSE
23	101	4.4	395	1	TRFR_MOUSE
24	101	4.4	481	1	Y389_MOUSE
25	100.5	4.3	514	1	COX1_MOUSE
26	100.5	4.3	530	1	COX1_MOUSE
27	100	4.3	386	1	NU4M_MOUSE
28	99.5	4.3	355	1	OPSG_MOUSE
29	99.5	4.3	371	1	CMU_MOUSE
30	99	4.3	373	1	CMU_MOUSE
31	99	4.3	527	1	NU2M_MOUSE
32	99	4.3	615	1	NUOL_MOUSE
33	98.5	4.3	394	1	TSGA_MOUSE

34	98.5	4.3	562	1	ATKA_MOUSE
35	98.5	4.3	889	1	TRK2_MOUSE
36	98	4.2	318	1	MAAR_MOUSE
37	98	4.2	337	1	TR3_MOUSE
38	98	4.2	446	1	NU4M_MOUSE
39	98	4.2	447	1	NU4M_MOUSE
40	98	4.2	525	1	COX1_MOUSE
41	98	4.2	663	1	CYOB_MOUSE
42	97.5	4.2	250	1	ATP6_MOUSE
43	97.5	4.2	485	1	YIDL_MOUSE
44	97.5	4.2	512	1	FLD1_MOUSE
45	97.5	4.2	582	1	Y596_MOUSE

ALIGNMENTS

RESULT 1
CYB_TRYBO
ID CYB_TRYBO STANDARD; PRT; 372 AA.
AC Q33568; Q35988; 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Cytochrome b.
GN MTCYB OR COB OR CYTB OR CYB.
OS Trypanoplasma borreli.
OG Mitochondrion.
OC Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Trypanoplasma.
OX NCBI_Taxid=5710;
RN [1]
RP SEQUENCE FROM N.A., AND RNA EDITING.
RC STRAIN=PG-0H;
RX MEDLINE=95059046; PubMed=7969154;
RA Maslov D.A., Simpson L.;
RT "RNA editing and mitochondrial genomic organization in the cryptobid
kinetoplastid protozoan Trypanoplasma borreli.";
RL Mol. Cell. Biol. 14:8174-8182(1994).
RN [2]
RP SEQUENCE FROM N.A., AND RNA EDITING.
RC STRAIN=TT-0H;
RA Lukes J., van den Burg J., Arts G.J., de Haan A., Oppendoes F.,
Sloof P., Benne R.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: Component of the ubiquinol-cytochrome c reductase
complex (complex III or cytochrome b-c1 complex), which is a
respiratory chain that generates an electrochemical potential
coupled to ATP synthesis (By similarity).
CC - COPOLYMER: Binds two heme groups non-covalently. Heme 1 (or BL or
b562) is low-potential and absorbs at about 562, and heme 2 (or BR
or b566) is high-potential and absorbs at about 566 (By
similarity).
CC - SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
cytochrome c1 and the Rieske protein (By similarity).
CC - RNA EDITING: Modified positions=Not applicable; Note:Some
positions are modified by RNA editing via nucleotide insertion or
deletion.
CC - SIMILARITY: Belongs to the cytochrome b family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; U1182; AA65017.1; ALT_SEQ.
CC EMBL; U1184; AA73455.1; -.
CC PIR; S52054; S52054.
CC InterPro: IPR005798; Cytochrome b6 C.
CC InterPro: IPR005797; Cytochrome b6 N.
CC Pfam; PF00032; cytochrome_b_c1.1.

DR Pfam: PF00033; Cytochrome b N¹; 1.
 DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
 DR PROSITE; PS00193; CYTOCHROME_B_OO; 1.
 KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
 KW Heme; RNA editing.
 FT METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 97 97 IRON 2 (HEME B562 AXIAL LIGAND).
 FT METAL 184 184 IRON 1 (HEME B562 AXIAL LIGAND).
 FT METAL 198 198 IRON 2 (HEME B562 AXIAL LIGAND).
 SQ SEQUENCE 372 AA; 44710 MW; DC4AF96804A0CDE6 CRC64;
 Query Match 5.4%; Score 126; DB 1; Length 372;
 Best Local Similarity 20.7%; Pred. No. 0.026;
 Matches 92; Conservative 64; Mismatches 124; Indels 164; Gaps 24;
 QY 10 LCIALALSLAMFTRFTTLLVHIFISLVILGILFVCGVLMWLYTD-----YTND 61
 DB 17 LCCLALSGDLFRVYGLGFMIGWIALOI-----LVGIC--LSWFFRCILPQWIFTL 68
 QY 62 LSIETDERENKVCVGFAYSTGI--TAVLVLFVLAKRKIKLVTELFOITNKAISSAP 119
 DB 69 IHKEFD-----LGIIRSLHIFISLVFLYFLHITIKVIFCLIPDS----- 111
 QY 120 FLIFQPLMTFAILIFFWVLMVAVLISLG-TAGAAQVMEGGQVEYK-----PLSGIRYMW 172
 DB 112 ---MLVWFGFGLIFILILAF--IGYTLPTCSMSVGLVFNINILATIPLSI----- 160
 QY 173 SYHLGLIWTSEI-----LACQMTIAG--VTCYFNKSKNDPDHETL 216
 DB 161 --YICWMCSEFINDFTLLKHSIHIFLPVLLFLIGAFVLAIF-----L 206
 QY 217 SLSLIL-----FFYHQTGVKVSFLISVARIPIRIIWMQNALKEQHGALSRYL 267
 DB 207 SSGDLDRPPFYERFFFLYLRDPLI--INICFLIYI----- 247
 QY 266 RCYCCFCWGLDKYHLHNONATTAINSTD-----FCTSAKAFKILSKNS 314
 DB 248 ---CIYW---FFVHEHSMIIVDTLTKSDKILPEWFFLSFGFKLVDPKF----- 292
 QY 315 SHFTSINCGDFIIFLKGVLVVCFTVF---GGLMAFVYNRAFOVAVPELLVAPEATLV 370
 DB 293 -----MGLFLF-----VLCFALFLFIANCILIFCYGSSLLM-MSLSLVLFY-LIC 337
 QY 371 AHSFLSVETVIDALFLCPADLE 394
 DB 338 VGGFLSLY-----VVLCPFLMME 355
 RESULT 2
 NT4M PROWI STANDARD; PRT; 523 AA.
 ID NT4M PROWI
 AC Q37617;
 DT 15-UTL-1998 (Rel. 36, Created)
 DT 15-UTL-1998 (Rel. 36, Last sequence update)
 DT 15-UTL-1998 (Rel. 36, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3).
 GN ND4 OR NAD4.
 OS Prototheca wickerhamii.
 OC Mitochondrion.
 OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
 OC Chlorellaceae; Prototheca.
 OX NCBI_TaxID=3111;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=263-11;
 RX MEDLINE=94180393; PubMed=8133522;
 RA Wolfe G., Plante I., Lang B.F., Kueck U., Burger G.,
 RT "Complete sequence of the mitochondrial DNA of the chlorophyte alga
 RL J. Mol. Biol. 237:75-86(1994).
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U02970; AAD12635.1; -
 DR PIR; T11916; T11916.
 DR InterPro; IPR003918; NADHub oxd4.
 DR InterPro; IPR001750; Oxidored_q1.
 DR Pfam; PF00361; oxidored_q1; 1.
 DR PRINTS; PRO1437; NOOXDRDASE4.
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
 SQ SEQUENCE 523 AA; 58945 MW; 2C558DD00935801D CRC64;
 Query Match 5.1%; Score 119; DB 1; Length 523;
 Best Local Similarity 18.1%; Pred. No. 0.12; Indels 118; Gaps 15;
 Matches 76; Conservative 63; Mismatches 163; Indels 118; Gaps 15;
 QY 11 CIALALSLAMFTRFTTLLVHIFISLVILGILFVCGVLMWLYTDNDLSIEDTER 70
 DB 151 CIAFLVETLMTLVFSLVDLLFYIPFESVLIMFIIIGV-WG-----SRER 196
 QY 71 ENMKCVLGFAYSTGITAVLVLFVLRKRIKLVTELFOITNKAISSAPPLFOPLMTA 130
 DB 197 KIRAAVOFLYTLFGSVLMALILLYPQTGIDIMLVISDFSETRCCLMLAFAPSPA 256
 QY 131 IL-----IFPW-----VLVAVLSIGTAG--AAQVMEGGQVEYKLSGI 168
 DB 257 VKYPMVPHVHMLBEAHEVAPTAGSVILAGILKLTGYGLRSIPLPFACIYFPL 313
 QY 169 RYMSYHLGLIWTSS-----EPLACQMTIAGAVVTCYFNKSKNDPDHETLSL 219
 DB 314 --IYMSVLAIVYTSCTTIROIIDLKILAVSSVAHMFITIGLFSQNT----- 359
 QY 220 SILFFYHQTGVKVSFLISVARIPIRIIWMQNALKEQHGALSRYLFCYCCFCWGLDK 279
 DB 360 ---QG--IGSLILMI-----SHGLVSPALFLCVGLVY----- 387
 QY 280 YLHLHNONAYTTAINGTDFCTSAKDAFKILSKNSHFPSINCF-----GDFLIIFG 331
 DB 388 -----DHRKRLIAYSGCGQWPIFALL--FVFFMANISLPGTSFGEFLVFIG 437
 QY 332 KVLVCEYVFGGLMANNRAFOVAVPELLVAFPYLYAH-SFLSVETVIDALFLCPA 390
 DB 438 SYQNSFVAPCAATGAVLGAAYALMLCNLI-----YGVSKDPFITWSDVVRREFEFA 492
 RESULT 3
 NT4M ACACA STANDARD; PRT; 497 AA.
 ID NT4M ACACA
 AC Q37375;
 DT 15-UTL-1998 (Rel. 36, Created)
 DT 15-UTL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3).
 GN ND4 OR NAD4.
 OS Acanthamoeba castellanii (Amoebea).
 OC Mitochondrion.
 OC Eukaryota; Acanthamoebidae; Acanthamoeba.
 OX NCBI_TaxID=5755;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 30010 / Neff.
 RX MEDLINE=95147275; PubMed=7644823;
 RA Burger G., Plante I., Loneragan K.M., Gray M.W.,
 RT "The mitochondrial DNA of the amoeboid protozoan, Acanthamoeba
 RL castellanii: complete sequence, gene content and genome
 CC organization".
 CC J. Mol. Biol. 245:522-537(1995).
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U12386; AAD11826.1; -
DR PIR: S53834; S53834.
DR InterPro: IPR003918; NADhub_oxred4.
DR InterPro: IPR001750; Oxidored_g1.
DR Pfam: PF00361; Oxidored_g1; 1.
DR PRINTS: PR01437; NDONDRDAS4.
DR Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
KW SEQUENCE 497 AA; 57768 MW; 9C0B637E65C0F3B5 CRC64;

Query Match 5.1%; Score 118; DB 1; Length 497;
Best Local Similarity 19.5%; Pred. No. 0.13;
Matches 90; Conservative 65; Mismatches 143; Indels 164; Gaps 22;

QY 11 CLIALALSLAMFPTFTTLVHFTSLVILG-----LLFVCGV 50
DB 140 CVLTL-----VEFYFFESILIPFELIGVWGSSQKRIHAYQLFFYTLGLSLMLGI 193
QY 51 LMLVYDYNDLSIEDTERENMKCVLGFALVSTGTVLVLVFLVAKRIKLVTELQI 110
DB 194 LVYSHIQTDRLVLYNV-----SFYKQILVMSFP-- 226
QY 111 TNKAISAPFLFQPLWTFALIFPW-----VLMVAVLSLGTGAQAQWEG 157
DB 227 -----FAFCVKVPLFPFHI-----WLPFAVEAPTVSGVILAGLVLTGLTGL----- 269
QY 156 GQVEYKPLSGIRYM-----NSYHLIGLWTFSEFLACQOMTIAQAVTCYFNSSKNDP 211
DB 270 -----LRFVPIPCFDATYFPLPVYT-----LCGLGIITTCSTTRQVLYK 310
QY 212 DHPILSLSLFPPYQGVVKSFLISVVRIPRIIWMVQNALKEQGHASRYLFCY 271
DB 311 KYIAVASYS-----HMSFVILGLFTSNIGIGSGVFLM-----SHGIVSSGLFCTG 358
QY 272 CCFNGLDK-----LHLNQAVTTTAINGTDFCTSKADAFKLKSSSHFTSINCFDFTI 328
DB 359 CVY-----DRKTRILYSGVSTMP--FSLC-----FILLISNISPPGSSSFGFELI 408
QY 329 FLG-----KVLVVCFTVFGGLMAFNRAFOVAVLPLLVAFAYVVAHSFL-----S 376
DB 409 LLGLPENNHFAALITPGLI-----LTAVYSIMLYNRIT--FNRLVYVYVYLRPSDFS 458
QY 377 VFEYVLDLFLFCFAVDLETNDGSS-----SEKPYFMDQEFISF 413
DB 459 KKEFVVGFI-FCFITILFGLKSGYIISLEAPLYV---YLSF 496

RESUL 4
GALR MOUSE STANDARD; PRT; 348 AA.
AC P56479;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE DE Galanin receptor type 1 (GALR1) (GALR1).
GN GALR1 OR GALR1 OR GALR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97415411, PubMed=9271210,
RA Wang S., He C., Maguire M.T., Clemmons A.L., Burtler R.E., Guzzi M.F.,
Strader C.D., Parker E.M., Bayne M.L.;

RT "Genomic organization and functional characterization of the mouse
RT Galr1 galanin receptor".
RT FEBS Lett. 411:225-230(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=98035873; PubMed=9367674;
RA Jacoby A.S., Webb G.C., Liu M.L., Kofler B., Hort Y.J., Fathi Z.,
RA Bottema C.D.K., Shine J., Iismaa T.P.;
RT "Structural organization of the mouse and human GALR1 galanin
RT receptor genes (galr and GALR) and chromosomal localization of the
RT mouse gene".
RT Genomics 45:496-508(1997).
CC - FUNCTION: Receptor for the hormone galanin. The activity of this
CC receptor is mediated by G proteins that inhibit adenylate cyclase
CC activity.
CC - SUBCELLULAR LOCATION: Integral membrane protein.
CC - PM: THREE CYSTEINE RESIDUES ARE FOUND IN THE C-TERMINUS, AT LEAST
CC ONE OF WHICH MAY BE PALMITOYLATED (BY SIMILARITY).
CC - SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Y15004; CA75237.1; -
DR EMBL: U90657; AAB87748.1; -
DR EMBL: U90655; AAB87748.1; JOINED.
DR EMBL: U90656; AAB87748.1; JOINED.
DR MGI: 1096364; Galr1.
DR InterPro: IPR00276; GPCR_Rhodopn.
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCRHOPOPSN.
DR PROSITE: PS00237; G_PROTEIN_RECPT_P1_1; 1.
DR PROSITE: PS00262; G_PROTEIN_RECPT_P2_1; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Lipoprotein; Palmitate.
FT DOMAIN 1 34
FT TRANSSEM 35 55
FT DOMAIN 56 70
FT TRANSSEM 71 91
FT DOMAIN 92 109
FT TRANSSEM 110 131
FT DOMAIN 132 151
FT TRANSSEM 152 172
FT DOMAIN 173 197
FT TRANSSEM 198 218
FT DOMAIN 219 247
FT TRANSSEM 248 268
FT DOMAIN 269 270
FT TRANSSEM 271 291
FT DOMAIN 292 348
FT CARBOHYD 7 7
FT CARBOHYD 12 12
FT CARBOHYD 182 182
FT DISULFID 106 186
FT LIPID 319 319
SQ SEQUENCE 348 AA; 39114 MW; 6F52D752BA19F9A CRC64;

Query Match 4.9%; Score 112.5; DB 1; Length 348;
Best Local Similarity 19.4%; Pred. No. 0.24;
Matches 83; Conservative 43; Mismatches 129; Indels 173; Gaps 17;

QY 33 VHFISLVILGLFPGCVLWLYDYNDLSIEDTERENMKCVLGFALVSTGTVLVLV 92
DB 31 VENFPTLVVFGILFAMGV-----GNSLVITV 57
QY 93 LIFVAKRIKLVTELQINKAISAPFLF-----QPLMTFAILT--FFWVLM 139

DB 58 IARSKPKRSTTNLF-ILNLSIADLAVLLFCIPQATVVALPTWVLAFCIKFIHYEFT 116
QY 140 VAVLSLGTAGAAYV-----MEGGVEYMPISGIRVMSYHILGLIMSEFLA 188
DB 117 VSNLVSIFTLAAMSDRYAIVHSKRSSSLKRSALAGVGTMA-----161
QY 189 COQMTTGAAVTCY---FNRSKNDP-----PDHPILSSILFPGYQTVVKSFLIS 238
DB 162 ---LSIIMASPVAYHQRLPHRDSNQPCWECWPNK-----LHKKAYVCTEVEG 207
QY 239 VVRIPRIIWMQNLMKEQHGALSRYLERCYCCGFWCLDKYLLHNLQNAATTATNIGTD 298
DB 208 -----YLLPLILITF-CYAKVNLHMK-----229
QY 239 FCTSAKDAFKILSKNSHFTSINCFGDFIIFGKVLVVCFTVFGGMAFNVRAPQVW-- 356
DB 230 -----LKNMSKKSSEASKKKT-----AQTVLVVVVVG--ISMVPHVYHMAE 270
QY 357 --AVPLLVAFAYLVAH-----SEISV-FEYVLDALFLCPAVDLETNDGS 399
DB 271 FGAFPLTPASFFFRITLHCLAVSNSSVNPDIYAFISENFRAYKQVFCVCHVDSPPSET 330
QY 400 SEKPYEMD 407
DB 331 KENSRMD 338

RESULT 5

NU4M MARPO STANDARD; PRT; 495 AA.

ID NU4M MARPO

AC P26848;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DE NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3).

OS NADH OR NADH.

OS Marchantia polymorpha (Liverwort).

OC Mitochondrion.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;

OC Marchantiopsida; Marchantiidae; Marchantiales; Marchantiineae;

OC Marchantiaceae; Marchantia.

OC NCBI_TaxID=3197;

OK NCBI_TaxID=3197;

RN NCBI_TaxID=3197;

RP SEQUENCE FROM N.A.

RX MEDLINE=92114051; PubMed=1731062;

RA Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Nozato N.,

RA Akashi K., Kanegae T., Ogura Y., Kohchi T., Ohyama K.;

RT "Gene organization deduced from the complete sequence of liverwort

Marchantia polymorpha mitochondrial DNA. A primitive form of plant

mitochondrial genome.";

RT J. Mol. Biol. 223:1-7(1992).

RL J. Mol. Biol. 223:1-7(1992).

RN NCBI_TaxID=3197;

RP SEQUENCE FROM N.A.

RX MEDLINE=93247547; PubMed=8483448;

RA Nozato N., Oda K., Yamato K., Ohta E., Takemura M., Akashi K.,

RA Fukushima H., Ohyama K.;

RT "Cotranscriptional expression of mitochondrial genes for subunits of

NADH dehydrogenase, nad5, nad4, nad2, in Marchantia polymorpha.";

RL Mol. Genet. 237:343-350(1993).

CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@sib-sib.ch).

CC EMBL; M68929; AAC09398.1; .

CC PIR; S25942; S25942.

CC InterPro; IPR003918; NADHnb oxrtd4.

DR Pfam; PF00361; oxidored_g1; 1.
DR PRINTS; PR01437; NUXORDRAS54.
KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 495 AA, 56311 MW, 0F75894D6C4AED4 CRC64;

Query Match 4.8%; Score 110; DB 1; Length 495;
Best Local Similarity 18.9%; Pred. No. 0.52;
Matches 78; Conservative 73; Mismatches 152; Indels 110; Gaps 16;

QY 12 ILALSLAMFFRRITTLVHIFSLVILG-----LRFVCGVL 51
DB 130 LIAVFCSLDLIRYFESVLIPMTITIGWGSRRKIRAAVQFLYTLMSGLFMLALIL 189
QY 52 WMLYDYVNDLSIELTERENMKCVLGAIVSGITAVLVILFVLRKRIKLTVELFOIT 111
DB 190 FIFQGTGTDLIQLTTR-----FSRRQILMTAFESF 224
QY 112 NKAISAPFLFQPLTFAILIFFWLVAVLSLTAG---AAQVEGGVEYKPLSG 167
DB 225 SVKVPVPAVIMVPEAHVAPTAGSVTLAIGILKGTGFLRFSSIPMPPEATLYFTP--- 281
QY 168 IRYMASVYHILGLTSEFLIACQMTIAGAVVTCYFNRSKNDPDPHPILSSILFEYHQ 227
DB 282 --FTYLSVIAIYTS--LTTRQIDLKTLIA-----YSSVA-----HM 316
QY 228 GTVVKGSPLISVVRIPRIITWMQNALKEQHGALSRYLERCYCCGFWCLDRY---LML 284
DB 317 NFVTIGMFSLNIGIGISILLML-----SHGLVSSALFLCVGALY---DHRKTRIVXY 366
QY 285 NQNAVTTTANGDFTSAKDAFKILSKNSHFTSINCFGDFIIFG---KVLVVCFTV 340
DB 367 YGGLVSTMPFSTIF-----LFTLANWSLPGTS-SFISGFLILVGAFORNSLVATLAA 419
QY 341 FGGLMAFNVRAPQVAVPLLVAFPA--YLVASHLSVFETVLDALFLCPAV 391
DB 420 LGWIL-----GAASLWLVNRVAVFGNFKPFIILKFSQLRNREVL---IFLPFV 465

RESULT 6

TLCC_RICPR

ID TLCC_RICPR

AC 092D67;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE ADP/ATP carrier protein 3 (ADP/ATP translocase 3).

GN TLCC OR TLG3 OR RP477.

OS Rickettsia prowazekii.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OC NCBI_TaxID=782;

OK NCBI_TaxID=782;

RN NCBI_TaxID=782;

RP SEQUENCE FROM N.A.

RX STRAIN=Madrid E;

RX MEDLINE=99039499; PubMed=9823893;

RA Anderson S.G.E., Zomrodipour A., Anderson J.O.,

RA Scheritz-Ponten T., Almqvist U.C.M., Podowski R.M., Naeslund A.K.,

RA Eriksson A.-S., Winkler H.H., Kurland C.G.;

RT "The genome sequence of Rickettsia prowazekii and the origin of

mitochondria.";

RL Nature 396:1133-1140(1998).

CC -1- FUNCTION: PROVIDES THE RICKETTSIAL CELL WITH HOST ATP IN

EXCHANGE FOR RICKETTSIAL ADP. THIS IS AN OBLIGATE EXCHANGE SYSTEM.

THIS ENERGY ACQUIRING ACTIVITY IS AN IMPORTANT COMPONENT OF

RICKETTSIAL PARASITISM.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- SIMILARITY: Belongs to the ADP/ATP translocase tlc family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@sib-sib.ch).

Db 207 -----YLLPLLICF-CYAKVNLHKK----- 228
 QY 299 FCTSANDAFILSKNSHPTISNCFGDFIIFLGKLVVCFVFGMAFNRAFOVW-- 356
 Db 229 -----LKMSSKSSASKKT-----AQTVLVVVVVG--ISMVPHVHLMAE 269
 QY 357 --AVELLVAFPAIVH-----SFLSV-FETVADALFLCPAVULETNDGS 399
 Db 270 FGAPLPDPASFPRITAHCLAYNSVNPITVAFLESEFRKAVKQVFCVNCSEPHDA 329
 QY 400 SEK 402
 Db 330 KEK 332

RESULT 8

VATI_BORBU STANDARD; PRT; 608 AA.
 ID VATI_BORBU
 AC 05118;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE V-type ATP synthase subunit I (EC 3.6.3.14) (V-type ATPase subunit I).
 GN ATP1 OR BB0091.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 CC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D., Hanson M.,
 RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Uetback T., Matthey L., McDonald L., Artlich P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RA "Genomic sequence of a Lyme disease spirochete, Borrelia
 burgdorferi.";
 RT Nature 390:580-586(1997).
 RL Nature 390:580-586(1997).

CC - FUNCTION: Produces ATP from ADP in the presence of a proton
 CC gradient across the membrane.
 CC - CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
 CC H(+) (out).
 CC - SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL, AE001122; AAC66486.1; -.
 DR PIR, C70111; C70111.
 DR TIGR, BB0091; -.
 DR Interpro, IPR002490; V_ATPase_subI1.
 DR Pfam, PF01496; V_ATPase_sub_a1.
 KM Hydroxylase; Hydrogen ion transport; Transmembrane; Complete proteome.
 SQ SEQUENCE 608 AA; 65121 MW; 4B430DF23A521A36 CRC64;

Query Match 4.7%; Score 108.5; DB 1; Length 608;
 Best Local Similarity 17.6%; Pred. No. 0.82;
 Matches 68; Conservative 81; Mismatches 135; Indels 103; Gaps 16;

QY 65 ELDTERENKVCYIGF-----AIVSTGTTATL-----LVTLVFLAKRI-KLTV 105
 Db 232 DNGTDWEDSYITGIVPASPQESLKNVAKAGFAAFADPDENDIITPIYIKKGIANTLAA 291

QY 106 ELFOITKKAISAPFLFQPLMTFALLIFENVLVANVLISGTGGAACWEGGVEKPL 165
 Db 292 PIFNI-----LEITIPYKRDV-SFIFMLFFVFPGMI-----GPAAGVI 332
 QY 166 SGIRYMSYHLIGIWTSEFIACQMT-----IGAVVTCVFNRSKDDP 211
 Db 333 -----FRLIGILSIFLKKGRPLTFPHGLFYLVSYSILGANTGWFG-SPLLE 383
 QY 212 DHPILSSISLFFHOGTVKGSFLISVRIPIRIIVYMNQ--ALKEQOH-GALSYLF 267
 Db 384 MEFILNSFKVSYLTERKNSVQNIIFICPSIGVQLISLHWNFFRQVEKEPHISIAQIGW 443
 QY 268 RGCYCCFMCIDKYLHLNQNAYTTAI-----NGTDPCTSKAKDFKI 309
 Db 444 LMCIVGLYLVNLNI-LSQSRFPMNVNVNVIYGVALLVVFQKQDSNF-----FKC 495
 QY 310 LSKN-----SSHPTSINCFGDFIIFLGKLVVCFVFGMAFNRAFOVWAVPLLV 364
 Db 496 ILKSPGIIHQPLTVSGFADIIISYIRLFV-----GLAGSISASFMTMSILKSS 548
 QY 365 FFAIVVMSFLSVFETVADALFLCPAV 391
 Db 549 NIGLIVAGIIVILFGHVLNIMLSLSV 575

RESULT 9

Y294_MYCGE STANDARD; PRT; 474 AA.
 ID Y294_MYCGE
 AC 049411; Q49351;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MG294.
 GN MG294.
 OS Mycoplasma genitalium.
 CC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=96026346; PubMed=756993;
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 RA Fritchmann J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
 RA Nguyen D.T., Uetback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
 RA Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RA "The minimal gene complement of Mycoplasma genitalium.";
 RT Science 270:397-403(1995).
 RL Science 270:397-403(1995).
 RN [2]
 RN SEQUENCE OF 86-197 FROM N.A.
 RP STRAIN=ATCC 33530 / G-37;
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=94075230; PubMed=8253680;
 RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III;
 RA "A survey of the Mycoplasma genitalium genome by using random
 RA sequencing.";
 RT J. Bacteriol. 175:7918-7930(1993).
 RL J. Bacteriol. 175:7918-7930(1993).
 CC - SIMILARITY: SOME, TO E.COLI YIHN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL, U39709; AAC71515.1; -.
 DR EMBL, U02243; AAA03399.1; -.
 DR PIR, B64232; E64232.
 DR TIGR, MG294; -.
 KM Hypothetical protein; Complete proteome.
 FT CONFLICT 191 191 S -> C (IN REF. 2).

```

FT 197 197 K -> N (IN REF. 2).
SQ SEQUENCE 474 AA; 52649 MW; 665273D90312D551 CRC64;
Query Match
Best Local Similarity 21.28; Score 108; DB 1; Length 474;
Matches 88; Conservative 55; Mismatches 132; Indels 140; Gaps 20;
QY 37 ISVITGLL--FVCGVLMW-----LYDYNDLSIELDTERENKCVIGFATVSTG 85
DB 22 LALIIIGAIIDVFYIAAPYVKNVVPNMLITGITED-----EVAIVTSISGVTLAQ 74
QY 86 ITAVLVLLIVLKRRIKLTVELFQITNKAIS-----SAPFLIFQPLMT- 129
DB 75 LPGGFLTRSSKRLFLSA-----ITTGALTFWLANILTKNOQSHDALFIOYCVIWM 130
QY 130 --AIIFFVVLWVAV-----LISLTAGAAQVMEGGQVEYPLSGIRYMSYHLI 177
DB 131 GISTSLIFMTPLWKLASQATKENQALGFGIGQAA-----NGIMGLIFIFI 177
QY 178 GLTWSEFILAQQMTIAGAVTCYFRSKNDPPHP-----ILSSSLIFPY 225
DB 178 ALITTSIF-----YPSGSENDSPFAVAFIIMLVITGFTVLFV 221
QY 226 HQTIVVKS--FLISYVR--IPRIIWMQNALKEQHGALSRYPFCYCCFCWCLDKYLL 282
DB 222 KEKPIKQSGQTLVSPKRNINQILVTLKMKMLLSFFLMGMYVQSTFA-----YILL 275
QY 283 HIANQNY-----TTTALNG--TDFTSADAF--KILSKNSHFTSINCEDPIIF- 329
DB 276 QMLQNAFLAPVVLVATVYIGIRYALRSAYVLCRLADCKS-----YILFLMTC 325
QY 330 --LGKLVVCFVTFVGGIAPFNVARQWAVPLLVAFAYIVASHSVETVL 382
DB 326 TVLGIVFVALFIIIGFVQINSAN-----ITLITF-----SSILYIFIGIL 365

RESULT 10
LMP2_EBV STANDARD; PRT; 497 AA.
ID LMP2_EBV
AC P13285;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE Gene terminal protein (Membrane protein LMP-2A/LMP-2B).
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_Taxid=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88283646; PubMed=2840285;
RA Laux G., Perricaudet M., Farrell P.J.;
RT "A spliced Epstein-Barr virus gene expressed in immortalized
RT lymphocytes is created by circularization of the linear viral
RT genome."
RL EMBO J. 7:769-774(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89095024; PubMed=2536113;
RA Sample J., Liebowitz D., Kieff E.;
RT "Two related Epstein-Barr virus membrane proteins are encoded by
RT separate genes."
RL J Virol. 63:933-937(1989).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tufnell P.S., Watnall B.G., Watson G.S., White O., Wilson R.,
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome."
RL Nature 310:207-211(1984).
CC -1- FUNCTION: MAY BE IMPORTANT FOR IMMORTALIZATION OF CELLS BY EBV.
CC -1- SUBCELLULAR LOCATION: Membrane (Probable).

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcement/
CC or send an email to license@ebi.ac.uk).
CC
DR EMBL; M24212; AAA45887.1; -.
DR EMBL; Y00835; CAA68762.1; -.
DR EMBL; V01555; -, NOT_ANNOTATED_CDS.
DR PIR; A30178; WMBELM.
KW Transmembrane.
FT CHAIN 1 497 MEMBRANE PROTEIN LMP-2A.
FT TRANSMEM 120 497 MEMBRANE PROTEIN LMP-2B.
FT TRANSMEM 122 141 POTENTIAL.
FT TRANSMEM 150 168 POTENTIAL.
FT TRANSMEM 178 198 POTENTIAL.
FT TRANSMEM 208 235 POTENTIAL.
FT TRANSMEM 242 259 POTENTIAL.
FT TRANSMEM 267 288 POTENTIAL.
FT TRANSMEM 300 316 POTENTIAL.
FT TRANSMEM 321 339 POTENTIAL.
FT TRANSMEM 352 373 POTENTIAL.
FT TRANSMEM 392 411 POTENTIAL.
FT TRANSMEM 419 443 POTENTIAL.
FT TRANSMEM 450 470 POTENTIAL.
SQ SEQUENCE 497 AA; 53011 MW; F4DC9B8C1FD83F1 CRC64;
Query Match
Best Local Similarity 4.68; Score 106; DB 1; Length 497;
Matches 72; Conservative 49; Mismatches 126; Indels 110; Gaps 13;
QY 6 TILGCIATLALSLAMMFTFRFTT---LIVHIFISLVIIIGLFFVGVLMWLYYDNTDL 62
DB 150 TVTATGTLASLLILAAVASSYAAQKLTPTVTLAVTTPAIC-LTWRIIDPPNSL 208
QY 63 SIEDTERENKCVLGFATVSTGIVAVLVLFVLRKRIKY-ELFOITNK 113
DB 209 LPAU-----LAAAGIGQIYVLMVLLIAYRRRR-RLTVCGGIMFLACVLVIYDA 261
QY 114 AISAAPFLFQPLMTFAFLIFFWLVAVAVLSGTGAQVMEGGQVEY--KLSGIRN 171
DB 262 VLQSPFLGAVTVVSMTLILAFVLMWSFGGIGTIGALTLIAALALIASLIGTLNL 321
QY 172 WSYHLGLTWSEFIACQMT-----IAGAVVTCYFN 205
DB 322 TTMFLMLMTLVVLVILICSSCSPKXILIALFLYALALLILASALIAQSLIDTNK 381
QY 206 SKNDPPHP-----ILSSSLIFTHQG-----TVKSGFLIS 238
DB 382 SLSTSEFIPFLFPMILLIYAGILFIALITLWSSGNRTYGPVFMCLGGLITWAGAVWL 441
QY 239 VVRIPRIIWMQNALKEQHGALSRYL-----FRCC-YCCFMC 277
DB 442 V-----MENTL-----LSAMILLAGFLIFLIGFALRGVIRCCRCYCYCL 481

RESULT 11
ST21_ARATH STANDARD; PRT; 677 AA.
ID ST21_ARATH
AC O04722;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Sulfate transporter 2.1 (AST68).
GN SUTR2.1 OR ATSG10180 OR T31P16-170.
OS Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_Taxid=3702;

```

RN SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND INDUCTION.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=98021500; PubMed=9380766;
 RA Takahashi H., Yamazaki M., Sasakura N., Watanabe A., Tenetsek T.,
 de Almeida Engler J., Engler G., Van Montagu M., Saito K.;
 "Regulation of sulfur assimilation in higher plants: a sulfate
 transporter induced in sulfate-starved roots plays a central role in
 Arabidopsis thaliana.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:11102-11107(1997).
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016721; PubMed=11130714;
 RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
 Miyajima N., Sasamoto S., Kikura T., Hosouchi T., Kawashima K.,
 Kohara M., Matsunoto M., Matsuno A., Muraki A., Nakayama S.,
 Nakazaki N., Naito K., Okumura S., Shino S., Takeuchi C., Wada T.,
 Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
 Huang E., Spiegel L., Gao J., O'Shaughnessy A., Preston R.,
 Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
 Storeking T., Pepin K., Spieth J., Sekhon N., Armstrong J., Becker M.,
 Belter E., Cordum H., Cordes J., Courtney L., Courtney W., Dante M.,
 Du H., Edwards U., Fryan J., Haakensen B., Lamar B., Latreille P.,
 Leonard S., Meyer R., Mulvaney E., Ozerky P., Riley A., Stromatt C.,
 Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedina N.,
 Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
 Kirchhoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,
 Mattiensen R., McComble W.R., Wilson R.K., Murphy G., Hancock I.,
 Volkmer G., Wandut R., Duesetnoet A., Stekema W., Pohl T.,
 Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,
 Langham S.-A., McCullagh B., Robben J., Grymoprez B., Zimmermann W.,
 Ranspenger U., Medler H., Balke K., Medler E., Peters S.,
 van Staveren M., Dirke W., Moolman P., Klein lankhorst R.,
 Weitzenecker T., Bothe G., Rose M., Hauf J., Bernier S., Hempel S.,
 Feilgenbach M., Lamberth S., Villarroel R., Giesen U., Ardiles W.,
 Berts O., Lemcke K., Kolesov G., Mayer K.F.X., Rüd S., Schoof H.,
 Schueller C., Zaccaria P., Mewes H.-W., Beyan W., Franz P.F.;
 "Sequence and analysis of chromosome 5 of the plant Arabidopsis
 thaliana.";
 RL Nature 408:823-826(2000).
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=22954850; PubMed=14593172;
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
 Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Phan P.K., Cheuk R.F.,
 Karlin-Newman G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
 Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
 Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Anstett Y.,
 Arakawa T., Bann J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
 Chao Q., Choy N., Ejima N., Goldsmith A.D., Gujral M., Hansen N.F.,
 Hayashizaki Y., Johnson-Hopson C., Huan V.W., Iida K., Karnes M.,
 Kawanishi S., Kosemura E., Ishida J., Jiang P.X., Jones T., Kawai J.,
 Kamitani A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
 Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
 Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 "Empirical analysis of transcriptional activity in the Arabidopsis
 genome.";
 RL Science 302:842-846(2003).
 RN FUNCTION, TISSUE SPECIFICITY, AND INDUCTION.
 RP MEDLINE=20387013; PubMed=10929111;
 RA Takahashi H., Watanabe-Takahashi A., Smith F.W., Blake-Kalif M.,
 Hawkesford M.J., Saito K.;
 "The roles of three functional sulphate transporters involved in
 uptake and translocation of sulphate in Arabidopsis thaliana.";
 RL Plant J. 23:171-182(2000).
 CC -1- FUNCTION: Low-affinity H(+)/sulfate cotransporter that may be
 involved in root-to-shoot translocation of sulfate. Plays a
 central role in the regulation of sulfate assimilation.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- TISSUE SPECIFICITY: Expressed in root cap, central cylinder of

CC roots and in vascular tissues of leaves.
 CC -1- INDUCTION: In roots by sulfate starvation or after selenate
 treatment.
 CC -1- SIMILARITY: Belongs to the SLC26A/Sulph transporter (TC 2.A.53)
 family.
 CC -1- SIMILARITY: Contains 1 STAS domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL: AB003590; BAA20084.1; -
 CC EMBL: AB003591; BAA20085.1; -
 CC EMBL: AL356332; CAB92059.1; -
 CC EMBL: AY062546; AL332624.1; -
 CC EMBL: AY093335; AAM13334.1; -
 CC FTR: T50022; T50022.
 CC InterPro: IPR002645; STAS.
 CC InterPro: IPR001902; Sulph_transp.
 CC Pfam: PF01740; STAS; 1.
 CC Pfam: PF00916; Sulfate_transp; 1.
 CC TIGRfam: TIGR00815; Sulp; 1.
 CC PROSITE: PS01130; SLC26A; 1.
 CC PROSITE: PS0801; STAS; 1.
 CC TRANSPORT: Symport; Sulfate transport; Transmembrane; Glycoprotein;
 KM Multigene family.
 FT DOMAIN 1 118
 FT TRANSMEM 119 139
 FT DOMAIN 140 141
 FT TRANSMEM 142 142
 FT TRANSMEM 143 143
 FT TRANSMEM 144 144
 FT TRANSMEM 145 145
 FT TRANSMEM 146 146
 FT TRANSMEM 147 147
 FT TRANSMEM 148 148
 FT TRANSMEM 149 149
 FT TRANSMEM 150 150
 FT TRANSMEM 151 151
 FT TRANSMEM 152 152
 FT TRANSMEM 153 153
 FT TRANSMEM 154 154
 FT TRANSMEM 155 155
 FT TRANSMEM 156 156
 FT TRANSMEM 157 157
 FT TRANSMEM 158 158
 FT TRANSMEM 159 159
 FT TRANSMEM 160 160
 FT TRANSMEM 161 161
 FT TRANSMEM 162 162
 FT TRANSMEM 163 163
 FT TRANSMEM 164 164
 FT TRANSMEM 165 165
 FT TRANSMEM 166 166
 FT TRANSMEM 167 167
 FT TRANSMEM 168 168
 FT TRANSMEM 169 169
 FT TRANSMEM 170 170
 FT TRANSMEM 171 171
 FT TRANSMEM 172 172
 FT TRANSMEM 173 173
 FT TRANSMEM 174 174
 FT TRANSMEM 175 175
 FT TRANSMEM 176 176
 FT TRANSMEM 177 177
 FT TRANSMEM 178 178
 FT TRANSMEM 179 179
 FT TRANSMEM 180 180
 FT TRANSMEM 181 181
 FT TRANSMEM 182 182
 FT TRANSMEM 183 183
 FT TRANSMEM 184 184
 FT TRANSMEM 185 185
 FT TRANSMEM 186 186
 FT TRANSMEM 187 187
 FT TRANSMEM 188 188
 FT TRANSMEM 189 189
 FT TRANSMEM 190 190
 FT TRANSMEM 191 191
 FT TRANSMEM 192 192
 FT TRANSMEM 193 193
 FT TRANSMEM 194 194
 FT TRANSMEM 195 195
 FT TRANSMEM 196 196
 FT TRANSMEM 197 197
 FT TRANSMEM 198 198
 FT TRANSMEM 199 199
 FT TRANSMEM 200 200
 FT TRANSMEM 201 201
 FT TRANSMEM 202 202
 FT TRANSMEM 203 203
 FT TRANSMEM 204 204
 FT TRANSMEM 205 205
 FT TRANSMEM 206 206
 FT TRANSMEM 207 207
 FT TRANSMEM 208 208
 FT TRANSMEM 209 209
 FT TRANSMEM 210 210
 FT TRANSMEM 211 211
 FT TRANSMEM 212 212
 FT TRANSMEM 213 213
 FT TRANSMEM 214 214
 FT TRANSMEM 215 215
 FT TRANSMEM 216 216
 FT TRANSMEM 217 217
 FT TRANSMEM 218 218
 FT TRANSMEM 219 219
 FT TRANSMEM 220 220
 FT TRANSMEM 221 221
 FT TRANSMEM 222 222
 FT TRANSMEM 223 223
 FT TRANSMEM 224 224
 FT TRANSMEM 225 225
 FT TRANSMEM 226 226
 FT TRANSMEM 227 227
 FT TRANSMEM 228 228
 FT TRANSMEM 229 229
 FT TRANSMEM 230 230
 FT TRANSMEM 231 231
 FT TRANSMEM 232 232
 FT TRANSMEM 233 233
 FT TRANSMEM 234 234
 FT TRANSMEM 235 235
 FT TRANSMEM 236 236
 FT TRANSMEM 237 237
 FT TRANSMEM 238 238
 FT TRANSMEM 239 239
 FT TRANSMEM 240 240
 FT TRANSMEM 241 241
 FT TRANSMEM 242 242
 FT TRANSMEM 243 243
 FT TRANSMEM 244 244
 FT TRANSMEM 245 245
 FT TRANSMEM 246 246
 FT TRANSMEM 247 247
 FT TRANSMEM 248 248
 FT TRANSMEM 249 249
 FT TRANSMEM 250 250
 FT TRANSMEM 251 251
 FT TRANSMEM 252 252
 FT TRANSMEM 253 253
 FT TRANSMEM 254 254
 FT TRANSMEM 255 255
 FT TRANSMEM 256 256
 FT TRANSMEM 257 257
 FT TRANSMEM 258 258
 FT TRANSMEM 259 259
 FT TRANSMEM 260 260
 FT TRANSMEM 261 261
 FT TRANSMEM 262 262
 FT TRANSMEM 263 263
 FT TRANSMEM 264 264
 FT TRANSMEM 265 265
 FT TRANSMEM 266 266
 FT TRANSMEM 267 267
 FT TRANSMEM 268 268
 FT TRANSMEM 269 269
 FT TRANSMEM 270 270
 FT TRANSMEM 271 271
 FT TRANSMEM 272 272
 FT TRANSMEM 273 273
 FT TRANSMEM 274 274
 FT TRANSMEM 275 275
 FT TRANSMEM 276 276
 FT TRANSMEM 277 277
 FT TRANSMEM 278 278
 FT TRANSMEM 279 279
 FT TRANSMEM 280 280
 FT TRANSMEM 281 281
 FT TRANSMEM 282 282
 FT TRANSMEM 283 283
 FT TRANSMEM 284 284
 FT TRANSMEM 285 285
 FT TRANSMEM 286 286
 FT TRANSMEM 287 287
 FT TRANSMEM 288 288
 FT TRANSMEM 289 289
 FT TRANSMEM 290 290
 FT TRANSMEM 291 291
 FT TRANSMEM 292 292
 FT TRANSMEM 293 293
 FT TRANSMEM 294 294
 FT TRANSMEM 295 295
 FT TRANSMEM 296 296
 FT TRANSMEM 297 297
 FT TRANSMEM 298 298
 FT TRANSMEM 299 299
 FT TRANSMEM 300 300
 FT TRANSMEM 301 301
 FT TRANSMEM 302 302
 FT TRANSMEM 303 303
 FT TRANSMEM 304 304
 FT TRANSMEM 305 305
 FT TRANSMEM 306 306
 FT TRANSMEM 307 307
 FT TRANSMEM 308 308
 FT TRANSMEM 309 309
 FT TRANSMEM 310 310
 FT TRANSMEM 311 311
 FT TRANSMEM 312 312
 FT TRANSMEM 313 313
 FT TRANSMEM 314 314
 FT TRANSMEM 315 315
 FT TRANSMEM 316 316
 FT TRANSMEM 317 317
 FT TRANSMEM 318 318
 FT TRANSMEM 319 319
 FT TRANSMEM 320 320
 FT TRANSMEM 321 321
 FT TRANSMEM 322 322
 FT TRANSMEM 323 323
 FT TRANSMEM 324 324
 FT TRANSMEM 325 325
 FT TRANSMEM 326 326
 FT TRANSMEM 327 327
 FT TRANSMEM 328 328
 FT TRANSMEM 329 329
 FT TRANSMEM 330 330
 FT TRANSMEM 331 331
 FT TRANSMEM 332 332
 FT TRANSMEM 333 333
 FT TRANSMEM 334 334
 FT TRANSMEM 335 335
 FT TRANSMEM 336 336
 FT TRANSMEM 337 337
 FT TRANSMEM 338 338
 FT TRANSMEM 339 339
 FT TRANSMEM 340 340
 FT TRANSMEM 341 341
 FT TRANSMEM 342 342
 FT TRANSMEM 343 343
 FT TRANSMEM 344 344
 FT TRANSMEM 345 345
 FT TRANSMEM 346 346
 FT TRANSMEM 347 347
 FT TRANSMEM 348 348
 FT TRANSMEM 349 349
 FT TRANSMEM 350 350
 FT TRANSMEM 351 351
 FT TRANSMEM 352 352
 FT TRANSMEM 353 353
 FT TRANSMEM 354 354
 FT TRANSMEM 355 355
 FT TRANSMEM 356 356
 FT TRANSMEM 357 357
 FT TRANSMEM 358 358
 FT TRANSMEM 359 359
 FT TRANSMEM 360 360
 FT TRANSMEM 361 361
 FT TRANSMEM 362 362
 FT TRANSMEM 363 363
 FT TRANSMEM 364 364
 FT TRANSMEM 365 365
 FT TRANSMEM 366 366
 FT TRANSMEM 367 367
 FT TRANSMEM 368 368
 FT TRANSMEM 369 369
 FT TRANSMEM 370 370
 FT TRANSMEM 371 371
 FT TRANSMEM 372 372
 FT TRANSMEM 373 373
 FT TRANSMEM 374 374
 FT TRANSMEM 375 375
 FT TRANSMEM 376 376
 FT TRANSMEM 377 377
 FT TRANSMEM 378 378
 FT TRANSMEM 379 379
 FT TRANSMEM 380 380
 FT TRANSMEM 381 381
 FT TRANSMEM 382 382
 FT TRANSMEM 383 383
 FT TRANSMEM 384 384
 FT TRANSMEM 385 385
 FT TRANSMEM 386 386
 FT TRANSMEM 387 387
 FT TRANSMEM 388 388
 FT TRANSMEM 389 389
 FT TRANSMEM 390 390
 FT TRANSMEM 391 391
 FT TRANSMEM 392 392
 FT TRANSMEM 393 393
 FT TRANSMEM 394 394
 FT TRANSMEM 395 395
 FT TRANSMEM 396 396
 FT TRANSMEM 397 397
 FT TRANSMEM 398 398
 FT TRANSMEM 399 399
 FT TRANSMEM 400 400
 FT TRANSMEM 401 401
 FT TRANSMEM 402 402
 FT TRANSMEM 403 403
 FT TRANSMEM 404 404
 FT TRANSMEM 405 405
 FT TRANSMEM 406 406
 FT TRANSMEM 407 407
 FT TRANSMEM 408 408
 FT TRANSMEM 409 409
 FT TRANSMEM 410 410
 FT TRANSMEM 411 411
 FT TRANSMEM 412 412
 FT TRANSMEM 413 413
 FT TRANSMEM 414 414
 FT TRANSMEM 415 415
 FT TRANSMEM 416 416
 FT TRANSMEM 417 417
 FT TRANSMEM 418 418
 FT TRANSMEM 419 419
 FT TRANSMEM 420 420
 FT TRANSMEM 421 421
 FT TRANSMEM 422 422
 FT TRANSMEM 423 423
 FT TRANSMEM 424 424
 FT TRANSMEM 425 425
 FT TRANSMEM 426 426
 FT TRANSMEM 427 427
 FT TRANSMEM 428 428
 FT TRANSMEM 429 429
 FT TRANSMEM 430 430
 FT TRANSMEM 431 431
 FT TRANSMEM 432 432
 FT TRANSMEM 433 433
 FT TRANSMEM 434 434
 FT TRANSMEM 435 435
 FT TRANSMEM 436 436
 FT TRANSMEM 437 437
 FT TRANSMEM 438 438
 FT TRANSMEM 439 439
 FT TRANSMEM 440 440
 FT TRANSMEM 441 441
 FT TRANSMEM 442 442
 FT TRANSMEM 443 443
 FT TRANSMEM 444 444
 FT TRANSMEM 445 445
 FT TRANSMEM 446 446
 FT TRANSMEM 447 447
 FT TRANSMEM 448 448
 FT TRANSMEM 449 449
 FT TRANSMEM 450 450
 FT TRANSMEM 451 451
 FT TRANSMEM 452 452
 FT TRANSMEM 453 453
 FT TRANSMEM 454 454
 FT TRANSMEM 455 455
 FT TRANSMEM 456 456
 FT TRANSMEM 457 457
 FT TRANSMEM 458 458
 FT TRANSMEM 459 459
 FT TRANSMEM 460 460
 FT TRANSMEM 461 461
 FT TRANSMEM 462 462
 FT TRANSMEM 463 463
 FT TRANSMEM 464 464
 FT TRANSMEM 465 465
 FT TRANSMEM 466 466
 FT TRANSMEM 467 467
 FT TRANSMEM 468 468
 FT TRANSMEM 469 469
 FT TRANSMEM 470 470
 FT TRANSMEM 471 471
 FT TRANSMEM 472 472
 FT TRANSMEM 473 473
 FT TRANSMEM 474 474
 FT TRANSMEM 475 475
 FT TRANSMEM 476 476
 FT TRANSMEM 477 477
 FT TRANSMEM 478 478
 FT TRANSMEM 479 479
 FT TRANSMEM 480 480
 FT TRANSMEM 481 481
 FT TRANSMEM 482 482
 FT TRANSMEM 483 483
 FT TRANSMEM 484 484
 FT TRANSMEM 485 485
 FT TRANSMEM 486 486
 FT TRANSMEM 487 487
 FT TRANSMEM 488 488
 FT TRANSMEM 489 489
 FT TRANSMEM 490 490
 FT TRANSMEM 491 491
 FT TRANSMEM 492 492
 FT TRANSMEM 493 493
 FT TRANSMEM 494 494
 FT TRANSMEM 495 495
 FT TRANSMEM 496 496
 FT TRANSMEM 497 497
 FT TRANSMEM 498 498
 FT TRANSMEM 499 499
 FT TRANSMEM 500 500
 FT TRANSMEM 501 501
 FT TRANSMEM 502 502
 FT TRANSMEM 503 503
 FT TRANSMEM 504 504
 FT TRANSMEM 505 505
 FT TRANSMEM 506 506
 FT TRANSMEM 507 507
 FT TRANSMEM 508 508
 FT TRANSMEM 509 509
 FT TRANSMEM 510 510
 FT TRANSMEM 511 511
 FT TRANSMEM 512 512
 FT TRANSMEM 513 513
 FT TRANSMEM 514 514
 FT TRANSMEM 515 515
 FT TRANSMEM 516 516
 FT TRANSMEM 517 517
 FT TRANSMEM 518 518
 FT TRANSMEM 519 519
 FT TRANSMEM 520 520
 FT TRANSMEM 521 521
 FT TRANSMEM 522 522
 FT TRANSMEM 523 523
 FT TRANSMEM 524 524
 FT TRANSMEM 525 525
 FT TRANSMEM 526 526
 FT TRANSMEM 527 527
 FT TRANSMEM 528 528
 FT TRANSMEM 529 529
 FT TRANSMEM 530 530
 FT TRANSMEM 531 531
 FT TRANSMEM 532 532
 FT TRANSMEM 533 533
 FT TRANSMEM 534 534
 FT TRANSMEM 535 535
 FT TRANSMEM 536 536
 FT TRANSMEM 537 537
 FT TRANSMEM 538 538
 FT TRANSMEM 539 539
 FT TRANSMEM 540 540
 FT TRANSMEM 541 541
 FT TRANSMEM 542 542
 FT TRANSMEM 543 543
 FT TRANSMEM 544 544
 FT TRANSMEM 545 545
 FT TRANSMEM 546 546
 FT TRANSMEM 547 547
 FT TRANSMEM 548 548
 FT TRANSMEM 549 549
 FT TRANSMEM 550 550
 FT TRANSMEM 551 551
 FT TRANSMEM 552 552
 FT TRANSMEM 553 553
 FT TRANSMEM 554 554
 FT TRANSMEM 555 555
 FT TRANSMEM 556 556
 FT TRANSMEM 557 557
 FT TRANSMEM 558 558
 FT TRANSMEM 559 559
 FT TRANSMEM 560 560
 FT TRANSMEM 561 561
 FT TRANSMEM 562 562
 FT TRANSMEM 563 563
 FT TRANSMEM 564 564
 FT TRANSMEM 565 565
 FT TRANSMEM 566 566
 FT TRANSMEM 567 567
 FT TRANSMEM 568 568
 FT TRANSMEM 569 569
 FT TRANSMEM 570 570
 FT TRANSMEM 571 571
 FT TRANSMEM 572 572
 FT TRANSMEM 573 573
 FT TRANSMEM 574 574
 FT TRANSMEM 575 575
 FT TRANSMEM 576 576
 FT TRANSMEM 577 577
 FT TRANSMEM 578 578
 FT TRANSMEM 579 579
 FT TRANSMEM 580 580
 FT TRANSMEM 581 581
 FT TRANSMEM 582 582
 FT TRANSMEM 583 583
 FT TRANSMEM 584 584
 FT TRANSMEM 585 585
 FT TRANSMEM 586 586
 FT TRANSMEM 587 587
 FT TRANSMEM 588 588
 FT TRANSMEM 589 589
 FT TRANSMEM 590 590
 FT TRANSMEM 591 591
 FT TRANSMEM 592 592
 FT TRANSMEM 593 593
 FT TRANSMEM 594 594
 FT TRANSMEM 595 595
 FT TRANSMEM 596 596
 FT TRANSMEM 597 597
 FT TRANSMEM 598 598
 FT TRANSMEM 599 599
 FT TRANSMEM 600 600
 FT TRANSMEM 601 601
 FT TRANSMEM 602 602
 FT TRANSMEM 603 603
 FT TRANSMEM 604 604
 FT TRANSMEM 605 605
 FT TRANSMEM 606 606
 FT TRANSMEM 607 607
 FT TRANSMEM 608 608
 FT TRANSMEM 609 609
 FT TRANSMEM 610 610
 FT TRANSMEM 611 611
 FT TRANSMEM 612 612
 FT TRANSMEM 613 613
 FT TRANSMEM 614 614
 FT TRANSMEM 615 615
 FT TRANSMEM 616 616
 FT TRANSMEM 617 617
 FT TRANSMEM 618 618
 FT TRANSMEM 619 619
 FT TRANSMEM 620 620
 FT TRANSMEM 621 621
 FT TRANSMEM 622 622
 FT TRANSMEM 623 623
 FT TRANSMEM 624 624
 FT TRANSMEM 625 625
 FT TRANSMEM 626 626
 FT TRANSMEM 627 627
 FT TRANSMEM 628 628
 FT TRANSMEM 629 629
 FT TRANSMEM 630 630
 FT TRANSMEM 631 631
 FT TRANSMEM 632 632
 FT TRANSMEM 633 633
 FT TRANSMEM 634 634
 FT TRANSMEM 635 635
 FT TRANSMEM 636 636
 FT TRANSMEM 637 637
 FT TRANSMEM 638 638
 FT TRANSMEM 639 639
 FT TRANSMEM 640 640
 FT TRANSMEM 641 641
 FT TRANSMEM 642 642
 FT TRANSMEM 643 643
 FT TRANSMEM 644 644
 FT TRANSMEM 645 645
 FT TRANSMEM 646 646
 FT TRANSMEM 647 647
 FT TRANSMEM 648 648
 FT TRANSMEM 649 649
 FT TRANSMEM 650 650
 FT TRANSMEM 651 651
 FT TRANSMEM 652 652
 FT TRANSMEM 653 653
 FT TRANSMEM 654 654
 FT TRANSMEM 655 655
 FT TRANSMEM 656 656
 FT TRANSMEM 657 657
 FT TRANSMEM 658 658
 FT TRANSMEM 659 659
 FT TRANSMEM 660 660
 FT TRANSMEM 661 661
 FT TRANSMEM 662 662
 FT TRANSMEM 663 663
 FT TRANSMEM 664 664
 FT TRANSMEM 665 665
 FT TRANSMEM 666 666
 FT TRANSMEM 667 667
 FT TRANSMEM 668 668
 FT TRANSMEM 669 669
 FT TRANSMEM 670 670
 FT TRANSMEM 671 671
 FT TRANSMEM 672 672
 FT TRANSMEM 673 673
 FT TRANSMEM 674 674
 FT TRANSMEM 675 675
 FT TRANSMEM 676 676
 FT TRANSMEM 677 677
 FT TRANSMEM 678 678
 FT TRANSMEM 679 679
 FT TRANSMEM 680 680
 FT TRANSMEM 681 681
 FT TRANSMEM 682 682
 FT TRANSMEM 683 683
 FT TRANSMEM 684 684
 FT TRANSMEM 685 685
 FT TRANSMEM 686 686
 FT TRANSMEM 687 687
 FT TRANSMEM 688 688
 FT TRANSMEM 689 689
 FT TRANSMEM 690 690
 FT TRANSMEM 691 691
 FT TRANSMEM 692 692
 FT TRANSMEM 693 693
 FT TRANSMEM 694 694
 FT TRANSMEM 695 695
 FT TRANSMEM 696 696
 FT TRANSMEM 697 697
 FT TRANSMEM 698 698
 FT TRANSMEM 699 699
 FT TRANSMEM 700 700
 FT TRANSMEM 701 701
 FT TRANSMEM 702 702
 FT TRANSMEM 703 703
 FT TRANSMEM 704 704
 FT TRANSMEM 705 705
 FT TRANSMEM 706 706
 FT TRANSMEM 707 707
 FT TRANSMEM 708 708
 FT TRANSMEM 709 709
 FT TRANSMEM 710 710
 FT TRANSMEM 711 711
 FT TRANSMEM 712 712
 FT TRANSMEM 713 713
 FT TRANSMEM 714 714
 FT TRANSMEM 715 715
 FT TRANSMEM 716 716
 FT TRANSMEM 717 717
 FT TRANSMEM 718 718
 FT TRANSMEM 719 719
 FT TRANSMEM 720 720
 FT TRANSMEM 721 721
 FT TRANSMEM 722 722
 FT TRANSMEM 723 723
 FT TRANSMEM 724 724
 FT TRANSMEM 725 725
 FT TRANSMEM 726 726
 FT TRANSMEM 727 727
 FT TRANSMEM 728 728
 FT TRANSMEM 729 729
 FT TRANSMEM 730 730
 FT TRANSMEM 731 731
 FT TRANSMEM 732 732
 FT TRANSMEM 733 733
 FT TRANSMEM 734 734
 FT TRANSMEM 735 735
 FT TRANSMEM 736 736
 FT TRANSMEM 737 737
 FT TRANSMEM 738 738
 FT TRANSMEM 739 739
 FT TRANSMEM 740 740
 FT TRANSMEM 741 741
 FT TRANSMEM 742 742
 FT TRANSMEM 743 743
 FT TRANSMEM 744 744
 FT TRANSMEM 745 745
 FT TRANSMEM 746 746
 FT TRANSMEM 747 747
 FT TRANSMEM 748 748
 FT TRANSMEM 749 749
 FT TRANSMEM 750 750
 FT TRANSMEM 751 751
 FT TRANSMEM 752 752
 FT TRANSMEM 753 753
 FT TRANSMEM 754 754
 FT TRANSMEM 755 755
 FT TRANSMEM

Db 280 H-----TFILGCSFLSFLITRTRIGKKYKLFMLPAIAPLAVVSTLWFLTKAD- 330

QY 211 PDHPI-----LSSLSI--LFFY--HOGTVKSGFLISVPRPIVW----- 248

Db 331 -EKGVTVAHKGKGLNPMSTIODDFTPHGGIAKGLIITAVALTBAVGRSAGIKG 389

QY 249 YWQNALKEQHGALSRYL--FRCCYCCFWCLDKYLLHNLQNAVTTAINGDTCSTAKDA 306

Db 390 YRLDGNKEMVAIGFANVLGSFTSCYAA-----TGSFSTRVAV--FAAGCETA 434

QY 307 FKLSKNSSHFHSINFGDPIIF--LGKLVVCFVFGGLMAFNTRARQWAVP----- 359

Db 435 MSNTVAWVVFVVALBCLTRLLYTPPAIILASLPLGLI--NINEALHMKVKDFEL 492

QY 360 LLLVAFPAVLVAHSPFVETVLDALFLCPA 390

Db 493 ALIGAFPGVL---FASVEIGLVAIVISFA 519

RESULT 12

Y294 MYCPN STANDARD; PRT; 475 AA.

AC P75366;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein MG294 homolog (A05_01f475).

GN MEN421 OR MP420.

OS Mycoplasma pneumoniae.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI_TaxID=2104;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 29342 / M129;

RX MEDLINE=97105885; PubMed=8948633;

RA Himmelfreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C., Hermann R.;

RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.";

RL Nucleic Acids Res. 24:4420-4449(1996).

CC -1- SIMILARITY: SOME, TO E.COLI Y1HN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AB000041; AAB96068.1; -

DR PIR: S73746; S73746.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 475 AA; 52731 MW; 5A1P1F0E4B9B86D CRC64;

Query Match 4.5%; Score 104.5; DB 1; Length 475;

Best Local Similarity 20.0%; Pred. No.1.3;

Matches 78; Conservative 56; Mismatches 153; Indels 103; Gaps 16;

QY 37 ISLVILGL--LFFCVGLVM-----LYDYTNDSLIELTERENKCVLGFALVSTG 85

Db 20 ILILITIGADVFIAPAYIKVVPNLHLYLGITED-----EVATLISIIQVYLAQ 72

QY 86 IFAVLAVLFFVAKRIKLT-----VELFOITN-----KAISAPFLPQPLWTF---A 130

Db 73 LPEGFLANFSSKILFLSEISGVTTFPLATNILLREOKNALFVQCVIWMGLMGTIS 132

QY 131 ILIFFVTLVAV-----LISLGTAGAAQVMEGGQVYKPLSGIRMSYHLGLIW 181

Db 133 TLIFFTPLWKLASQATQENQALFGIIOGAA-----NGV--WGFIFFIL- 175

QY 182 TSEFIACQOMTLAGAVVTCYFNRSKNDPDHPI-----LSSLSILFFHYQGTVV 231

Db 176 -----ALITRAVAVPAGSESSANNPAPAIYAFIIGGMLVTFGLVFPVEKPIE 226

QY 232 K-----GSFLISVVAIPRIIWMQNALKEQHGALSRYLFRCCYCCFCWCLDKYLLHNLQ 287

Db 227 KYDSHTSLTKAKKNFEOILITLKNWKLMLISFFLMGKTVQSTFA-----YLLQMQN 280

QY 288 AYTTAINGTDF-----CTSANDAFKILSKNSHFTSINCFDPIIFLGLYLVCFV 340

Db 281 AFLAPVLIGTVLGGVTRVYVRSVSV--LGRADKRSTYLFMLCTGTGILFVLMFIL 338

QY 341 FGLMARNRPAFYWAVPLLVAFPAVLV 370

Db 339 LG-----FGVGGQQNVYALITVSAIYL 362

RESULT 13

NPT1 RAT STANDARD; PRT; 465 AA.

AC Q62795;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Renal sodium-dependent phosphate transport protein 1 (Sodium/phosphate cotransporter 1) (Na(+)/Pi cotransporter 1) (Renal sodium-phosphate transport protein 1) (Renal Na(+)-dependent phosphate cotransporter 1).

GN Slc127A1 OR NPT1.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=97021433; PubMed=8867793;

RA Li H., Xie Z.;

RT "Molecular cloning of two rat Na+/Pi cotransporters: evidence for differential tissue expression of transcripts.";

RL Cell. Mol. Biol. Res. 41:451-460(1995).

CC -1- FUNCTION: Important for the resorption of phosphate by the kidney. May be involved in actively transporting phosphate into cells via Na(+)-cotransport in the renal brush border membrane.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- SIMILARITY: Belongs to the sodium/phosphate cotransporter family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: U28504; AAC52487.1; -

DR InterPro: IPR007114; MFS.

DR InterPro: IPR004745; Pi cotransp.

DR InterPro: IPR005829; Sub cotransporter.

DR Pfam: PF00083; sugar_tr_1.

DR TIGRfams: TIGR00894; ZAO114euk; 1.

DR PROSITE: PS00850; MFS; 1.

KW Transport; Symport; Sodium transport; Transmembrane; Glycoprotein.

CC TRANSMEM 79

FT TRANSMEM 117

FT TRANSMEM 137

FT TRANSMEM 176

FT TRANSMEM 199

FT TRANSMEM 260

FT TRANSMEM 280

FT TRANSMEM 299

FT TRANSMEM 319

FT TRANSMEM 337

FT TRANSMEM 357

FT TRANSMEM 363

FT TRANSMEM 383

FT TRANSMEM 399

FT TRANSMEM 419

FT TRANSMEM 431

FT TRANSMEM 451

FT CARBOHD 39

FT CARBOHD 47

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 56 56 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 465 AA; 51350 MW; 288BFB8D5C0AC52 CRC64;

Query Match 4.4%; Score 103; DB 1; Length 465;
Best Local Similarity 21.1%; Pred. No. 1.6;
Matches 71; Conservative 39; Mismatches 95; Indels 132; Gaps 17;

CY 36 FLSVITLLFVCGVLMVLYDYNDLSIEDTERENKCVLGFAVISTGTAVLILF 95
DB 181 FIALVSG--FICDILGHPMVFYI-----IGVGVCLFWLTLF 219
CY 96 -----VLRKIKLTVEFQITNKAISSAP--FLFQPLMTFAILFFVWLWVAVL 143
DB 220 DDPNHPVMSSEKQYITSSLMQVHSGRSLPIKAMLSPLW--AII----- 266
CY 144 LSLGRGAAGVMEGGQVYKPLSGIRVMSYHLGLTSEFLLACQMTAGAVTCYF 203
DB 267 -----LNSPAFIWSNNLL-VYTPYTFI-----STYLVH 293
CY 204 NRSKNDPDPHPLLSLSILFFVHQGVV--KGSFLI-----SVRIPIIWMQNALKE 256
DB 294 NVRENG-----LSSLPLVLAIVIGIVAGMSDLSKIRSVAVAKLTTL----- 341
CY 257 QQHGALSYLPRCCYCCWCWCDKULHINQNAVTT--TAINGT--DPTSAKDAFKI 309
DB 342 --GIFCPVIFVVC-----LLVSYNFSYTVIFLTANSTLSFSFGQLNALDI 388
CY 310 LSKNSHPTSINCFDPIIFLPGVLYVCFVFGGMA 346
DB 389 APR-----YGFLL-KAVTALIGIFGLIS 411

RESULT 14
ID FLOI CRIGR STANDARD; PRT; 518 AA.

AC P42557;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Folate transporter 1 (Folate carrier protein) (Methotrexate uptake protein).
GN STC19A1.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
CC NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=94164933; PubMed=8119923;
RA Williams F.M.R., Murray R.C., Underhill T.M., Flintoff W.F.;
RT "Isolation of a hamster cDNA clone coding for a function involved in methotrexate uptake."
RL J. Biol. Chem. 269:5810-5816(1994).
CC -1- FUNCTION: TRANSPORTER FOR THE INTAKE OF FOLATE. INVOLVED IN METHOTREXATE UPTAKE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE STC19A FAMILY OF TRANSPORTERS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; U03031; AAC52138.1; -.
DR PIR; A53207; A53207.
DR InterPro; IPR002666; Folate carrier.
DR Pfam; PF01770; Folate carrier; 1.
DR Trifams; TIGR00806; Ftc; 1.

KM Folate-binding; Transport; Transmembrane; Glycoprotein.
FT TRANSEM 28 48 POTENTIAL.
FT TRANSEM 65 85 POTENTIAL.
FT TRANSEM 91 111 POTENTIAL.
FT TRANSEM 122 142 POTENTIAL.
FT TRANSEM 156 176 POTENTIAL.
FT TRANSEM 180 202 POTENTIAL.
FT TRANSEM 274 294 POTENTIAL.
FT TRANSEM 304 324 POTENTIAL.
FT TRANSEM 331 351 POTENTIAL.
FT TRANSEM 362 382 POTENTIAL.
FT TRANSEM 398 418 POTENTIAL.
FT TRANSEM 432 452 POTENTIAL.
SQ SEQUENCE 518 AA; 58611 MW; 0B18267A134FC5A CRC64;

Query Match 4.4%; Score 103; DB 1; Length 518;
Best Local Similarity 19.3%; Pred. No. 1.8;
Matches 76; Conservative 68; Mismatches 133; Indels 116; Gaps 20;

CY 67 DTERENKCVLGFAVISTGTAVL-----LVLIFFVLRKIKLTVEFQITNKAISSAP-- 119
DB 19 DDELKMRCLV-FYLCFQPMQALRPGSEFIPYILQ--NFTIE--QVTEIIFVLYS 73
CY 120 -----FLFQPLMTFAILFFVWLWVAVLISLGTAGA-AQVMEGGQVYKPL 165
DB 74 HLAIVPIPLTDVLRKYLILLOCLISFMCV-W--LLILLGTSVVMQIME--VFYSVT 127
CY 166 SGIRVMSYHLGLTSEF-----ILACQMTAGAVTCFVNSKNDPDPH 214
DB 128 MAARTAYSSYISLVRPSRYOMASYRAAVLGVFTSSVLAQVIMPEQSSONS----- 182
CY 215 ILSLSILFFVHQGVVKGSPFISVRIPIIWMQNAL-----KEQOH----- 259
DB 183 -----NMLNYSILGFIIFSLGLSLFLKRPKSLFNRSAVYKALPCELDQHPGAPB 237
CY 260 -GALSRYLPRC-----CYCCWCWCDK-----YLVH-----LN 285
DB 238 PGLTELVGSGNSFLVCLSELVGNLRQPHVRLCWMVFNRSAGYVLIYVHVLSID 297
CY 286 QNAVYTTAINGTDFCTSAKDAFKILSKNSHPTSINCFDPIIFLPGVLYVCFVFGGIM 345
DB 298 KVLNNGAVDASTLSLTSF-----SAGVYKIR-----WALMSKLYIASVIALQGLV 347
CY 346 AFVYRFAQVAVPPLVLAFFAYLV--AHSTL 375
DB 348 FCMYVHVYVTV-VHKIWLVMYTVYLFAGAYQFL 379

RESULT 15

ID CTAA_BACSU STRAND; PRT; 306 AA.
AC P12946;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome Aa3 controlling protein.
GN CTAA OR BSU14870.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CC NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89359135; PubMed=2549006;
RA Mueller J.P., Gaber H.W.;
RT "Isolation and sequence of ctat, a gene required for cytochrome a33 biosynthesis and sporulation in Bacillus subtilis."
RL J. Bacteriol. 171:4967-4978(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=169;
RT Bertero M., Presciani E., Glaser P., Richou A., Darchin A.;
RL "Bacillus subtilis chromosomal region downstream nprE"; Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.

[3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=168:
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertolo M.G., Bessières P., Bolotin A., Borchert S.,
 RA Borstis R., Bourasier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capiano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Ertien K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Paro T., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Prescan E., Pujic P., Purnelle B., Radoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadala Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccini E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassaretto A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzneger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256 (1997).
 CC -! FUNCTION: CTA is REQUIRED FOR CYTOCHROME AA3 BIOSYNTHESIS AND
 CC INDEPENDENTLY FOR SPOULATION
 CC -! SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M23915; AAA50254.1; -
 DR EMBL; Z98682; CAB11340.1; -
 DR EMBL; Z99111; CAB13360.1; -
 DR PIR; A33960; A33960.
 DR Subtilist; BG10213; ctaa.
 DR InterPro; IPR003780; COX15_Ctaa.
 DR Pfam; PF02628; COX15_Ctaa; 1.
 KW Sporulation; Membrane; Complete proteome.
 SQ SEQUENCE 306 AA; 34085 MW; D47FE9041851CDD CRC64;

Query Match 4.4%; Score 102.5; DB 1; Length 306;
 Best Local Similarity 21.7%; Pred.No.1.2; Mismatches 53; Gaps 10;
 Matches 53; Conservative 39; Indels 87; Gaps 65; Indels 10;

QY 9 GUCIALALSLAMFTFRTITLVH-IFISLVILGLFVCGVLMWLYDYNDISILD 67
 DB 65 GISII-LVYSILA-FMSWKITPIPRETPFLAINSIIFLQALV-----GALAVVFG 114
 QY 68 TERENMKCVLGFATVSTGTAVLVLFILAKRIKLYELFPI----- 110
 DB 115 SNALIMALHFGISLISFASVLITLILFEADRSVRLVPLQIGKMOFHMIGILIVSYI 174
 QY 111 -----TKKAISAPFLFOPV-----W-----TPAILIFFWLWVAVLL 144
 DB 175 VVYTGAYVHTSSSLACPNVPLCSPLNNGLPQGFHEWQMGHRAAALLLFWIIVAAVHA 234

QY 145 SLGTAGAQQVMEG-----QVEYKPLSGIRYMSYHILGLIMTSEFLACQQTIGAVY 139
 DB 235 ITSYKOKQQLFWGWSICLIFFITLQALSGIMIVISELALGFLAHSFFIAC-----LFG--V 288
 QY 200 TCYP 203
 DB 289 LCVF 292
 Search completed: May 27, 2004, 16:36:02
 Job time : 20 secs

Tue Jun 1 09:37:58 2004

us-10-063-686-32.ra1

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:34:25 ; Search time 22 Seconds
(without alignments)
1044.253 Million cell updates/sec

Title: US-10-063-686-32

Perfect score: 2315
Sequence: 1 MSGRDTLGLCTIALALSLA.....QDDKSLRNEGTLEQATVR 445

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/1aa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTUS COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	207	8.9	116	US-08-702-344-21	Sequence 21, Appl
2	161.5	7.0	148	US-09-247-155-166	Sequence 166, Appl
3	112.5	4.9	348	US-08-513-974B-46	Sequence 46, Appl
4	112.5	4.9	348	US-08-993-088A-10	Sequence 10, Appl
5	112.5	4.9	348	US-08-993-424B-10	Sequence 10, Appl
6	112.5	4.9	348	US-08-540-650B-2	Sequence 2, Appl
7	112.5	4.9	348	US-09-595-549-5	Sequence 5, Appl
8	112.5	4.9	348	US-09-461-436B-46	Sequence 46, Appl
9	112.5	4.9	348	US-09-603-680-10	Sequence 10, Appl
10	110.5	4.8	348	US-08-513-974B-342	Sequence 342, Appl
11	108.5	4.7	346	US-09-199-737-5	Sequence 5, Appl
12	108.5	4.7	346	US-08-993-088A-3	Sequence 3, Appl
13	108.5	4.7	346	US-08-993-424B-3	Sequence 3, Appl
14	108.5	4.7	346	US-09-058-333A-5	Sequence 5, Appl
15	108.5	4.7	346	US-09-595-549-6	Sequence 6, Appl
16	108.5	4.7	346	US-09-603-680-3	Sequence 3, Appl
17	108.5	4.7	346	US-08-899-112B-30	Sequence 30, Appl
18	107.5	4.6	349	US-08-513-974B-343	Sequence 343, Appl
19	107	4.6	345	US-08-981-700A-5	Sequence 5, Appl
20	106	4.6	497	US-09-718-693A-1	Sequence 1, Appl
21	102.5	4.4	395	US-08-900-230-5	Sequence 5, Appl
22	102	4.4	378	US-09-045-583-5	Sequence 5, Appl
23	102	4.4	378	US-09-534-185-5	Sequence 5, Appl
24	102	4.4	394	US-09-134-000C-3732	Sequence 3732, Ap
25	102	4.4	446	US-07-626-618A-21	Sequence 21, Appl
26	102	4.4	446	US-08-333-977-21	Sequence 21, Appl
27	102	4.4	446	US-07-969-267B-4	Sequence 4, Appl

28	102	4.4	446	US-09-168-510-4	Sequence 4, Appl
29	102	4.4	606	US-09-107-532A-4683	Sequence 4683, Ap
30	99.5	4.3	597	US-09-328-352-4703	Sequence 4703, Ap
31	99	4.3	940	US-09-328-352-8165	Sequence 8165, Ap
32	98.5	4.3	889	US-08-118-101A-4	Sequence 4, Appl
33	98	4.2	318	US-08-345-696-25	Sequence 25, Appl
34	98	4.2	318	US-08-233-009-25	Sequence 25, Appl
35	98	4.2	318	US-08-560-231-25	Sequence 25, Appl
36	98	4.2	318	US-09-080-704A-25	Sequence 25, Appl
37	98	4.2	318	US-09-800-274-6	Sequence 6, Appl
38	98	4.2	486	US-09-134-001C-3593	Sequence 3593, Ap
39	98	4.2	663	US-09-711-164-332	Sequence 332, Ap
40	97.5	4.2	374	US-09-102-710B-3	Sequence 3, Appl
41	97	4.2	335	US-09-387-699-4	Sequence 4, Appl
42	97	4.2	335	US-09-641-259B-4	Sequence 4, Appl
43	97	4.2	454	US-09-252-991A-30013	Sequence 30013, A
44	97	4.2	613	US-09-170-486D-212	Sequence 212, App
45	96.5	4.2	454	US-09-134-001C-3501	Sequence 3501, Ap

ALIGNMENTS

RESULT 1
US-08-702-344-21
Sequence 21, Application US/08702344

Patent No. 5723315
GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavallee, Edward
APPLICANT: Racine, Lisa
APPLICANT: Treacy, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESSES:
ADDRESSES: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/702,344
FILING DATE:

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 116 amino acids
TYPE: amino acid
STRANDNESS:

TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-702-344-21

Query Match 8.9%; Score 207; DB 1; Length 116;
Best Local Similarity 38.3%; Pred. No. 2e-14;
Matches 46; Conservative 19; Mismatches 29; Indels 26; Gaps 4;

Tue Jun 1 09:37:58 2004

us-10-063-686-32.ra1

Page 2

```

QY      269  CCYCCFACGCDKRYLHLHNOVAYTTALNGDFPCSSAKDAFKIKSSHSPTINCGPEII 328
Db      3    CLKCCFPCGJEKFKFLNBNAYIMIALYGNFPCSSARNAFELLMRIIIKVAIVDKYTBELF 62

QY      329  FLKRLVYCTVAGVGMAF-----NRYARQAVNAPVLLVAEPATLVN 371
Db      3    LKGLKLLVGSV---GLIAFFFFTHRRIRYQDPAFLNAY----YV-VIILVIVGSYILA 113

```

```

RESULT 2
US-09-247-155-166
Sequence 166, Application US/09247155A
Patent No. 6312822
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, Jean-Baptiste
APPLICANT: Ducrest, Americ
APPLICANT: Bouqueleret, Lydie
TITLE OF INVENTION: Complementary DNAs
FILE REFERENCE: GENSET, 021A
CURRENT APPLICATION NUMBER: US/09/247,155A
CURRENT FILING DATE: 1999-02-09
EARLIER APPLICATION NUMBER: 60/074,121
EARLIER FILING DATE: 1998-02-09
EARLIER APPLICATION NUMBER: 60/081,563
EARLIER FILING DATE: 1998-04-13
EARLIER APPLICATION NUMBER: 60/096,116
EARLIER FILING DATE: 1998-08-10
EARLIER APPLICATION NUMBER: 60/099,273
EARLIER FILING DATE: 1998-10-04
NUMBER OF SEQ. ID NOS: 162
SOFTWARE: Patent.pm
SEQ ID NO 166
LENGTH: 148
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SIGNAL
LOCATION: -48...-1
NAME/KEY: UNSURE
LOCATION: 32,100
OTHER INFORMATION: Xaa = any one of the twenty amino acids
US-09-247-155-166

```

```

Query Match      7.0%; Score 161.5; DB 4; Length 148;
Best Local Similarity 32.2%; Pred. No. 2.2e-09;
Matches 47; Conservative 20; Mismatches 50; Indels 29; Gaps 6

QY      293 AINGDPTCTSADPAFKILSKNSHSTINCGRDEPIIFLKGVLV-----CFTVFGLMA 346
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db       3 AIGKRFCSAKNAFLMRNIVRVVLDKXTDLLFPGLLVGVGIVGSFFPFGRIIP 62
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||

QY      347 -----FNINRAFOYNAVPELLIVAFAYLVAHSFLSVETVDALPLCPAVDLERN 396
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db       63 GKQDFKSPHLNV-----YW-LPKMTSILGAIVIASGFSGVCMDTLPFCEDLE-- 114
      |||::||::||::||::||::||::||::||::||::||::||::||::||::||

QY      397 DGSSEKPY---FMODEFLSFVRSRK 419
      ::||::||::||::||::||::||::||::||::||::||::||::||::||
Db       115 --RTAPWATLAHVQELLKLGGKKE 138
      ::||::||::||::||::||::||::||::||::||::||::||::||::||

```

RESULT 3
US-08-513-974B-46
; Sequence 46, Application US/08513974B
; Patent No. 6114139
; GENERAL INFORMATION:
; ADDRESS: 111
; ADDRESS: 111

```

1 TITLE OF INVENTION: PRODUCTION, AND USE THEREOF
2 NUMBER OF SEQUENCES: 380
3 CORRESPONDENCE ADDRESS:
4 ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
5 STREET: 130 Water Street
6 CITY: Boston
7 STATE: MA
8 COUNTRY: USA
9 ZIP: 02109
10
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE: floppy disk
13 COMPUTER: IBM PC compatible
14 OPERATING SYSTEM: PC-DOS/MS-DOS
15 SOFTWARE: PatentIn Release #1.0, Version #1.30
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER: US/08/513,974B
18 FILING DATE: 14-SEP-1995
19 CLASSIFICATION: 536
20 PRIOR APPLICATION DATA:
21 APPLICATION NUMBER: PCT/JP95/01599
22 FILING DATE: 10-AUG-1995
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: JP 7-093989
25 FILING DATE: 19-AUG-1995
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: JP 7-057186
28 FILING DATE: 16-MAR-1995
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: JP 7-007177
31 FILING DATE: 20-JAN-1995
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: JP 6-326611
34 FILING DATE: 28-DEC-1994
35 PRIOR APPLICATION DATA:
36 APPLICATION NUMBER: JP 6-270017
37 FILING DATE: 02-NOV-1994
38 PRIOR APPLICATION DATA:
39 APPLICATION NUMBER: JP 6-236357
40 FILING DATE: 30-SEP-1994
41 PRIOR APPLICATION DATA:
42 APPLICATION NUMBER: JP 6-236356
43 FILING DATE: 30-SEP-1994
44 PRIOR APPLICATION DATA:
45 APPLICATION NUMBER: JP 6-189274
46 FILING DATE: 11-AUG-1994
47 PRIOR APPLICATION DATA:
48 APPLICATION NUMBER: JP 6-189273
49 FILING DATE: 11-AUG-1945
50 PRIOR APPLICATION DATA:
51 APPLICATION NUMBER: JP 6-189272
52 FILING DATE: 11-AUG-1994
53 ATTORNEY/AGENT INFORMATION:
54 NAME: Resnick, David S.
55 REGISTRATION NUMBER: 34,235
56 REFERENCE/DOCKET NUMBER: 45753
57 TELECOMMUNICATION INFORMATION:
58 TELEPHONE: 617-523-3400
59 TELEFAX: 617-523-6440
60 INFORMATION FOR SEQ ID NO: 46:
61 SEQUENCE CHARACTERISTICS:
62 LENGTH: 348 amino acids
63 TYPE: amino acid
64 STRANDEDNESS:
65 TOPOLOGY: linear
66 MOLECULE TYPE: peptide
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

DB 31 VENTITLVFGLIFAMGVL-----GNSLVITV 57
QY 93 LIFVLRRIKLTVELFQITNKAISSAPLLE-----QPLMTFAILL---FWVLW 139
DB 58 LARSKPGKPRSTYLF-ILNLSIADLAVLFCIPFQATVVALPTWVGAFCKEIHVFT 116
QY 140 VAVLISLGTAGAACV-----MEGQVEYKPLSGIRYMSYHLIGIMTSEFLA 188
DB 117 VSMVLSITFLAASVDYRVAIVHSRRSSSLRVSNNALGVGFIVA-----161
QY 189 COQMTIAGAVTCTY---FNRSKNDP-----PDHPIISSLSILFFVHGQTVVKGSLIS 238
DB 162 ---LSIMASPVAVHQRLFHRDSNQTFCMEQWPNK-----LHKAYVVCFTVFG 207
QY 239 VVRIPIRIWYMONALKKEQHGALSRYLFRCCYCFWCLDKYLLHNLONATTTAINGTD 298
DB 208 -----YLLPILLICF-CYAKVLNHLHK-----229
QY 299 FCTSAKDAFKILSKNSHFTSINCFGDFIIFLGKVLVVCFTVFGMAFVYNAFOVW--356
DB 230 -----LKNMSKSEASKKT-----AQTVLVVVVFG--ISWLPHVHVMAB 270
QY 357 --AVPLLVAFPAVLVAH-----SFLSV-FETVLDALFLCPAVDLETNDGS 399
DB 271 FGAFLPTPASFFRITAHCLAVSNSVNPITIVAFLESENFRKAYKVCFKCHVCESPRSET 330
QY 400 SEKPYFMD 407
DB 331 KENKSRMD 338

RESULT 4
US-08-993-088A-10
; Sequence 10, Application US/08993088A
; Patent No. 6287855
; GENERAL INFORMATION:
; APPLICANT: Tan, Carina
; APPLICANT: Sullivan, Kathleen
; TITLE OF INVENTION: GALANIN RECEPTOR GALR2 AND
; TITLE OF INVENTION: NUCLEOTIDES ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, 126 E. Lincoln Ave.
; CITY: Rahway
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065-0900
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,088A
; FILING DATE: 18-DEC-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/033,851
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Heber, Sheldon O.
; REGISTRATION NUMBER: 38,179
; REFERENCE/DOCKET NUMBER: 19846
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732-594-1958
; TELEFAX: 732-594-4720
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-993-088A-10
Query Match 4.9%; Score 112.5; DB 3; Length 348;
Best Local Similarity 19.4%; Pred. No. 0.0013;
Matches 83; Conservative 43; Mismatches 129; Indels 173; Gaps 17;
QY 33 VHIISLVILGLFVCGVLMVLYDYDNLSDIELTERENKCVLGFALVSTGITALVLLV 92
DB 31 VENTITLVFGLIFAMGVL-----GNSLVITV 57
QY 93 LIFVLRRIKLTVELFQITNKAISSAPLLE-----QPLMTFAILL---FWVLW 139
DB 58 LARSKPGKPRSTYLF-ILNLSIADLAVLFCIPFQATVVALPTWVGAFCKEIHVFT 116
QY 140 VAVLISLGTAGAACV-----MEGQVEYKPLSGIRYMSYHLIGIMTSEFLA 188
DB 117 VSMVLSITFLAASVDYRVAIVHSRRSSSLRVSNNALGVGFIVA-----161
QY 189 COQMTIAGAVTCTY---FNRSKNDP-----PDHPIISSLSILFFVHGQTVVKGSLIS 238
DB 162 ---LSIMASPVAVHQRLFHRDSNQTFCMEQWPNK-----LHKAYVVCFTVFG 207
QY 239 VVRIPIRIWYMONALKKEQHGALSRYLFRCCYCFWCLDKYLLHNLONATTTAINGTD 298
DB 208 -----YLLPILLICF-CYAKVLNHLHK-----229
QY 299 FCTSAKDAFKILSKNSHFTSINCFGDFIIFLGKVLVVCFTVFGMAFVYNAFOVW--356
DB 230 -----LKNMSKSEASKKT-----AQTVLVVVVFG--ISWLPHVHVMAB 270
QY 357 --AVPLLVAFPAVLVAH-----SFLSV-FETVLDALFLCPAVDLETNDGS 399
DB 271 FGAFLPTPASFFRITAHCLAVSNSVNPITIVAFLESENFRKAYKVCFKCHVCESPRSET 330
QY 400 SEKPYFMD 407
DB 331 KENKSRMD 338

RESULT 5
US-08-993-424B-10
; Sequence 10, Application US/08993424B
; Patent No. 6337206
; GENERAL INFORMATION:
; APPLICANT: Tan, Carina
; APPLICANT: Kolakowski, Lee F., Jr.
; TITLE OF INVENTION: MOUSE GALANIN RECEPTOR GALR2 AND
; TITLE OF INVENTION: NUCLEOTIDES ENCODING SAME
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, 126 E. Lincoln Ave.
; CITY: Rahway
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065-0900
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,424B
; FILING DATE: 18-DEC-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/033,851
; FILING DATE: 27-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Heber, Sheldon O.
; REGISTRATION NUMBER: 38,179

Tue Jun 1 09:37:58 2004

us-10-063-686-32.fat

Page 4

REFERENCE/DOCKET NUMBER: 19846NP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 732-594-1958
TELEFAX: 732-594-4720
TELEX:
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-993-424B-10

Query Match 4.9%; Score 112.5; DB 4; Length 348;
Best Local Similarity 19.4%; Pred. No. 0.0013;
Matches 83; Conservative 43; Mismatches 129; Indels 173; Gaps 17;

QY 33 VHFISLVILGLIFVCGVLMVLYDYTDNLSIELTERENMKCVLGFAIVSTGTTAVLV 92
DB 31 VENFTLVVFGILFAMGVL-----GNSLVITV 57
QY 93 LIFVLRKRIKLTVELFOITNKAISSAPFLF-----OPLMTFAILI--FFVVLW 139
DB 58 LARSKGKPRSTTNLF-ILNLSIADLAVILFCIPQATVVALPTVVLGAFICKFIHYEFT 116
QY 140 VAVILSLGTAGAAQV-----MEGGQVEYKPLSGIRYMSYHLIGLIWTSFILA 188
DB 117 VSMIVSIFTLAAMSDRYVALVHSRRSSSLRVSRNALGVGFIMA-----161
QY 189 CQOMTIGAVVTCY---FNRSKNDP-----PDHPISSLSILFFYHOGTVKGSFLIS 238
DB 162 ---LSIAMSPPVAHYQRLEFHRDSNOTFCWEQPNK-----LHKKAYVCTFVFG 207
QY 239 VVRIPIRIYVMQNALKEQHGALSRVLFRCYCCFCWCLDKYLHLNQNATTTAINGTD 298
DB 208 -----YLLPLLLICF-CYAKVLNHLHKK-----229
QY 239 FCTSAKDAFKILSKNSHFTSINCFDPIIFLGKVLVCTVFGGLMAFNTRAFQVW-- 356
DB 230 -----LKNMSKSEASKKT-----AQTVLVVVVFG--ISMLPHVVLHMAE 270
QY 357 --AVPLLVAFPAIVAH-----SFLSV-FETVLDALFLCFVADLETNDGS 399
DB 271 FGAFLTPASFFPRITAHCLAVSNSVNPPIYAFLENFRAYKQVFKCHVCDSPRSET 330
QY 400 SEKPYFMD 407
DB 331 KENKSRMD 338

RESULT 6
US-08-540-650B-2
Sequence 2, Application US/08540650B
Patent No. 6399325
GENERAL INFORMATION:
APPLICANT: HINDA, Shuji
APPLICANT: FUJII, Ryo
APPLICANT: FUKUSUMI, Shoji
APPLICANT: OHYAKI, Tetsuya
APPLICANT: HOSOTA, Masaki
APPLICANT: OHGI, Kazuhiro
APPLICANT: ONDA, Haruo
TITLE OF INVENTION: GALANIN RECEPTOR PROTEIN, PRODUCTION AND USE THEREOF
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
STREET: 130 WATER STREET
CITY: BOSTON
STATE: MA
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/540,650B
FILING DATE: 11-OCT-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 7-134412
FILING DATE: 31-MAY-1995
APPLICATION NUMBER: 6-326610
FILING DATE: 28-DEC-1994
APPLICATION NUMBER: 6-247599
FILING DATE: 13-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: RESNICK, DAVID S
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 45901
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 348
TYPE: Amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-540-650B-2

Query Match 4.9%; Score 112.5; DB 4; Length 348;
Best Local Similarity 19.4%; Pred. No. 0.0013;
Matches 83; Conservative 43; Mismatches 129; Indels 173; Gaps 17;

QY 33 VHFISLVILGLIFVCGVLMVLYDYTDNLSIELTERENMKCVLGFAIVSTGTTAVLV 92
DB 31 VENFTLVVFGILFAMGVL-----GNSLVITV 57
QY 93 LIFVLRKRIKLTVELFOITNKAISSAPFLF-----OPLMTFAILI--FFVVLW 139
DB 58 LARSKGKPRSTTNLF-ILNLSIADLAVILFCIPQATVVALPTVVLGAFICKFIHYEFT 116
QY 140 VAVILSLGTAGAAQV-----MEGGQVEYKPLSGIRYMSYHLIGLIWTSFILA 188
DB 117 VSMIVSIFTLAAMSDRYVALVHSRRSSSLRVSRNALGVGFIMA-----161
QY 189 CQOMTIGAVVTCY---FNRSKNDP-----PDHPISSLSILFFYHOGTVKGSFLIS 238
DB 162 ---LSIAMSPPVAHYQRLEFHRDSNOTFCWEQPNK-----LHKKAYVCTFVFG 207
QY 239 VVRIPIRIYVMQNALKEQHGALSRVLFRCYCCFCWCLDKYLHLNQNATTTAINGTD 298
DB 208 -----YLLPLLLICF-CYAKVLNHLHKK-----229
QY 239 FCTSAKDAFKILSKNSHFTSINCFDPIIFLGKVLVCTVFGGLMAFNTRAFQVW-- 356
DB 230 -----LKNMSKSEASKKT-----AQTVLVVVVFG--ISMLPHVVLHMAE 270
QY 357 --AVPLLVAFPAIVAH-----SFLSV-FETVLDALFLCFVADLETNDGS 399
DB 271 FGAFLTPASFFPRITAHCLAVSNSVNPPIYAFLENFRAYKQVFKCHVCDSPRSET 330
QY 400 SEKPYFMD 407
DB 331 KENKSRMD 338

RESULT 7
US-09-595-549-5
Sequence 5, Application US/09595549
Patent No. 651827
GENERAL INFORMATION:
APPLICANT: Howard, Andrew D.

APPLICANT: Cascieri, Margaret A.
APPLICANT: Smith, Roy G.
APPLICANT: Sullivan, Kathleen A.
APPLICANT: Tan, Carla
APPLICANT: Van der Ploeg, Leonardus H. T.
APPLICANT: Lynch, Kevin R.
TITLE OF INVENTION: GALANIN RECEPTOR GALR3 AND NUCLEOTIDES
FILE REFERENCE: 20148PCA
CURRENT APPLICATION NUMBER: US/09/595,549
PRIOR FILING DATE: 2000-06-16
PRIOR APPLICATION NUMBER: US98/26612
PRIOR FILING DATE: 1998-12-17
PRIOR APPLICATION NUMBER: 60/069,725
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 348
TYPE: PRT
ORGANISM: mouse
US-09-595-549-5

Query Match 4.9%; Score 112.5; DB 4; Length 348;
Best Local Similarity 19.4%; Pred. No. 0.0013;

Matches 83; Conservative 43; Mismatches 129; Indels 173; Gaps 17;

QY 33 VHIISLVILGLFVCGVLMWLYDYNDISIELDERENMKCVLGFALVSTGTAVALV 92
DB 31 VENNITLVVGLIFAMGVL-----GNSLIVTV 57
QY 93 LIFVLRRKIKLTVELFOITNKATISAPFLF-----QPLWTFALII---FFWVLM 139
DB 58 LARSKPKRSTNLF--ILNLSLADIALYLLFCIPQATYVALFTWVGAFICFHYFFPT 116
QY 140 VAVILSLGTAGAAY-----MEGQVEYKPLSGIRYMSYHLGLIWTSEFLA 188
DB 117 VSMVLVSIFTLAAMSVDRYVALIVHSRRSSSLRVSRNALGVGFMA----- 161
QY 189 CQOMTIGAVTVCY---FNRKNDP-----PDHPILSLSLIFFYHGTIVKSGFLIS 238
DB 162 ---LSIAVMSPAVYHQRLFRHDSNQTFCEWQWPNK-----LHKAVVCTFVFG 207
QY 239 VVRIPIIVYMNALKEQOQHALSRYLFRCCYCCFCWCDKXILHLNQAVYTTAINGTD 298
DB 208 -----YLLPDLILIC-CYAKYLNHLHK----- 229
QY 299 FCTSAADAFKILSKSSSHFTSINCFGDFIIFGKVLVCTVFGGLMAFNVRARQVW-- 356
DB 230 -----LKNMSKSEASKKT-----AQTLLVVVVEG--ISWLPFHVHLWAE 270
QY 357 --AVPLLVAFPAVLVAH-----SPLSV--FEYVLDALFLCFANDLFTNGS 399
DB 271 FGAFFLTPASFPRIRIYHCLAYSNSVNPITYAFISENFRKAYKQVFKCHVCDSPRSET 330
QY 400 SEKPYFMD 407
DB 331 KENKSRMD 338

RESULT 8
US-09-461-436B-46
Sequence 46, Application US/09461436B
Patent No. 6538107
GENERAL INFORMATION:
APPLICANT: Shuji Hinuma
Yasunaki Ito
Ryo Fujii
TITLE OF INVENTION: G Protein Coupled Receptor Protein,
Production, And Use Thereof
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edwards & Angel, LLP

STREET: 101 Federal Street
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02209
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/461,436B
FILING DATE: 14-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/513,974
FILING DATE: 14-SEP-1995
APPLICATION NUMBER: PCT/JP95/01599
FILING DATE: 10-AUG-1995
APPLICATION NUMBER: 7-093989
FILING DATE: 19-APR-1995
APPLICATION NUMBER: 7-057186
FILING DATE: 16-MAR-1995
APPLICATION NUMBER: 7-007177
FILING DATE: 20-JAN-1995
APPLICATION NUMBER: 6-326611
FILING DATE: 28-DEC-1994
APPLICATION NUMBER: 6-270017
FILING DATE: 02-NOV-1994
APPLICATION NUMBER: 6-236357
FILING DATE: 30-SEP-1994
APPLICATION NUMBER: 6-236356
FILING DATE: 30-SEP-1994
APPLICATION NUMBER: 6-189274
FILING DATE: 11-AUG-1994
APPLICATION NUMBER: 6-189273
FILING DATE: 11-AUG-1994
APPLICATION NUMBER: 6-189272
FILING DATE: 11-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: CONLIN, DAVID G.
REGISTRATION NUMBER: <Unknown>
REFERENCE/DOCKET NUMBER: 45753 DIV2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-439-4444
TELEFAX: 617-439-4170
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-461-436B-46
Query Match 4.9%; Score 112.5; DB 4; Length 348;
Best Local Similarity 19.4%; Pred. No. 0.0013;
Matches 83; Conservative 43; Mismatches 129; Indels 173; Gaps 17;

QY 33 VHIISLVILGLFVCGVLMWLYDYNDISIELDERENMKCVLGFALVSTGTAVALV 92
DB 31 VENNITLVVGLIFAMGVL-----GNSLIVTV 57
QY 93 LIFVLRRKIKLTVELFOITNKATISAPFLF-----QPLWTFALII---FFWVLM 139
DB 58 LARSKPKRSTNLF--ILNLSLADIALYLLFCIPQATYVALFTWVGAFICFHYFFPT 116
QY 140 VAVILSLGTAGAAY-----MEGQVEYKPLSGIRYMSYHLGLIWTSEFLA 188
DB 117 VSMVLVSIFTLAAMSVDRYVALIVHSRRSSSLRVSRNALGVGFMA----- 161
QY 189 CQOMTIGAVTVCY---FNRKNDP-----PDHPILSLSLIFFYHGTIVKSGFLIS 238

Db 162 ---LSTAMSPVAYHQRLFHRDSNOTFCMEQPNK-----LHKAYVCTVFG 207
QY 239 VVARIPRIWMQNALKEQHGALSRYLFRCCYCCFCLDKYLLHNONAYTTAINGTD 298
Db 208 -----YLLPLLICF-CYAXVNLHKK----- 229
QY 239 FCTSAKDAFKILSKNSHFTSINCQDPIIFLGKYLVCFTVFGGLMAFNRAQVW-- 356
Db 230 -----LKNSSKSEASKKT-----AQTLYVVVVFQ--ISWLPHHVHLMAB 270
QY 357 --AVPLLVAFPAVYVAH-----SFLSV-FETVLDALFLCPAVDLETDG 399
Db 271 FGAFPLTPASFFPRITAHCLAYSNSVNPIYAFISENFRKAYKQVFKCHVCDSPRSET 330
QY 400 SEKPYFMD 407
Db 331 KENKSRMD 338

RESULT 9
US-09-603-680-10
Sequence 10, Application US/09603680
Patent No. 6544753
GENERAL INFORMATION:
APPLICANT: Sullivan, Carina
TITLE OF INVENTION: GALANIN RECEPTOR GAIK2 AND
NUCLEOTIDES ENCODING SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000, 126 E. Lincoln Ave.
CITY: Rahway
STATE: NJ
COUNTRY: USA
ZIP: 07065-0900
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/603,680
FILING DATE: 26-Jun-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/033,851
FILING DATE: 27-DEC-1996
APPLICATION NUMBER: 08/993,088
FILING DATE: 18-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 19846 CA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 732-594-1958
TELEFAX: 732-594-4720
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-603-680-10

Query Match 4.9%; Score 112.5; DB 4; Length 348;
Best Local Similarity 19.4%; Pired. No. 0.0013;
Matches 83; Conservative 43; Mismatches 129; Indels 173; Gaps 17;

QY 33 VAFISLVLSGLFPGCVLWLYDYTDLSIEDTERENKCVLGFALVSTGIANVLLV 92
Db 31 VENFITLVVFGILIFMAGVL-----GNSLVITV 57
QY 93 LIFVLAKRIKILVTELEFQITNKAISSAPFLF-----OPLWTFAILI--FEVYIWM 139
Db 58 IASKKGRKPRSTNLF-LNLSIADAYLLFCIPQAVYALPTVWLGAFICKFIHYFTT 116
QY 140 VAVLSLGTGAQAQV-----MEGGQVEYEPPLSIRKMSYHLGLIWTSEFILA 188
Db 117 VSMVLSIFTLAAMSVDRYVAIVHSRSSLRVSRNALGVGFIMW----- 161
QY 189 CQOMTAGAVVTCY-----FNRSKNDP-----PDHILSLSLFPHYQGVVKGSPFIS 238
Db 162 ---LSTAMSPVAYHQRLFHRDSNOTFCMEQPNK-----LHKAYVCTVFG 207
QY 239 VVARIPRIWMQNALKEQHGALSRYLFRCCYCCFCLDKYLLHNONAYTTAINGTD 298
Db 208 -----YLLPLLICF-CYAXVNLHKK----- 229
QY 239 FCTSAKDAFKILSKNSHFTSINCQDPIIFLGKYLVCFTVFGGLMAFNRAQVW-- 356
Db 230 -----LKNSSKSEASKKT-----AQTLYVVVVFQ--ISWLPHHVHLMAB 270
QY 357 --AVPLLVAFPAVYVAH-----SFLSV-FETVLDALFLCPAVDLETDG 399
Db 271 FGAFPLTPASFFPRITAHCLAYSNSVNPIYAFISENFRKAYKQVFKCHVCDSPRSET 330
QY 400 SEKPYFMD 407
Db 331 KENKSRMD 338

RESULT 10
US-08-513-974B-342
Sequence 342, Application US/08513974B
Patent No. 6114139
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Hosoya, Masaki
APPLICANT: Fujii, Ryo
APPLICANT: Ohtaki, Tetsuya
APPLICANT: Fukusumi, Shoji
APPLICANT: Ohgi, Kazuniro
TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,
NUCLEOTIDES ENCODING G PROTEIN COUPLED RECEPTOR PROTEIN,
NUMBER OF SEQUENCES: 380
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/513,974B
FILING DATE: 14-SEP-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP95/01599
FILING DATE: 10-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-093989
FILING DATE: 19-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-057186
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 7-007177
FILING DATE: 20-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-326611
FILING DATE: 28-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-270017
FILING DATE: 02-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236357
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236356
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189274
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189273
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189272
FILING DATE: 11-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S.
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 45753
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 342:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-513-974B-342

Query Match 4.8%; Score 110.5; DB 3; Length 348;

Best Local Similarity 19.4%; Pred. No. 0.0022; Matches 83; Conservative 42; Mismatches 130; Indels 173; Gaps 17;

33 VHFISVLIGLFGVGLMWLYDYDNDLSIEDTEREMKCVLGFAIVSTGTAVALLV 92
31 VENFTLVVFGILFAMGVL-----GNSLVITV 57
93 LIFVLRKRIKLTVELFOITNKAISSAPFLF-----OPLMTFAILL---FFWVLW 139
58 LARSKDGNPRSTNPLF--ILNLSIADLAYLLFCIPQATVVALPTVWLGAFICKFIHYFT 116
140 VAVLSLGTAGAAY-----MEGQVEYKPLSGIRYKMSYHLGLMTSEPIIA 188
117 VSMVLSIFTLAASVDRYVAIVHSRRSSSLRVSRNALLGVGFTWA----- 161
189 COQMTIAGAVTTCY-----FNRSKNDP-----PDHPIISLSILFFYHOGTVVKGSPFLIS 238
162 ---LSIMASPVAYHQRLEFHRDSONQFCWEQWPNK-----LHKAVVVCFTVFG 207
239 VRIPIRIIWMQNALKEQHGALSRVLPCCYCCFWCLDKTLHLNQNAYTTTAINGTD 298
208 -----YLLPDLILICF-CYAKVNLHLKK----- 229
299 FCTSAKDAFKLSKNSHFTSINCFDFTIFLGKVLVVCFTVGGIAMAENYRAFOVW-- 356
230 -----LKNMSKKSSEASKKKT-----AQTVLVVVVFG--ISWLPHHVYHIMAE 270
357 --AVPLLVAFFAYLVAN-----SFLSV-FETVLDALFLCPAVDLETNDGS 399
271 FGAFPLTPASFFFRITAHCLAYNSSVNPIIYAFISENFRKAYKQVFKCHVCDSEPRSET 330
400 SEKPYEMD 407
331 KENKSRMD 338

RESULT 11
US-09-199-737-5
Sequence 5, Application US/09199737A
Patent No. 6287788
GENERAL INFORMATION:
APPLICANT: Bard, Jonathan A.
APPLICANT: Borowsky, Beth
APPLICANT: Smith, Kelli E.
APPLICANT: Branchek, Theresa A.
APPLICANT: Gerald, Christophe P.G.
TITLE OF INVENTION: DNA Encoding Galanin GALR3 Receptors And Uses Thereof
FILE REFERENCE: 52241-D-PCT-US
CURRENT FILING DATE: 1998-11-25
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.0 - beta
SEQ ID NO 5
LENGTH: 346
TYPE: PRT
ORGANISM: Rat
US-09-199-737-5

Query Match 4.7%; Score 108.5; DB 3; Length 346;
Best Local Similarity 19.1%; Pred. No. 0.0036; Matches 81; Conservative 45; Mismatches 124; Indels 173; Gaps 17;

33 VHFISVLIGLFGVGLMWLYDYDNDLSIEDTEREMKCVLGFAIVSTGTAVALLV 92
30 VENFTLVVFGILFAMGVL-----GNSLVITV 56
93 LIFVLRKRIKLTVELFOITNKAISSAPFLF-----OPLMTFAILL---FFWVLW 139
57 LARSKDGNPRSTNPLF--ILNLSIADLAYLLFCIPQATVVALPTVWLGAFICKFIHYFT 115
140 VAVLSLGTAGAAY-----MEGQVEYKPLSGIRYKMSYHLGLMTSEPIIA 188
116 VSMVLSIFTLAASVDRYVAIVHSRRSSSLRVSRNALLGVGFTWA----- 160
189 COQMTIAGAVTTCY-----FNRSKNDP-----PDHPIISLSILFFYHOGTVVKGSPFLIS 238
161 ---LSIMASPVAYHQRLEFHRDSONQFCWEHWPNQ-----LHKAVVVCFTVFG 206
239 VRIPIRIIWMQNALKEQHGALSRVLPCCYCCFWCLDKTLHLNQNAYTTTAINGTD 298
207 -----YLLPDLILICF-CYAKVNLHLKK----- 228
299 FCTSAKDAFKLSKNSHFTSINCFDFTIFLGKVLVVCFTVGGIAMAENYRAFOVW-- 356
229 -----LKNMSKKSSEASKKKT-----AQTVLVVVVFG--ISWLPHHVYHIMAE 269
357 --AVPLLVAFFAYLVAN-----SFLSV-FETVLDALFLCPAVDLETNDGS 399
270 FGAFPLTPASFFFRITAHCLAYNSSVNPIIYAFISENFRKAYKQVFKCHVCDSEPRSET 332
400 SEK 402
330 KEK 332

RESULT 12
US-08-993-088A-3
Sequence 3, Application US/08993088A
Patent No. 6287855
GENERAL INFORMATION:
APPLICANT: Tan, Carina
APPLICANT: Sullivan, Kathleen
TITLE OF INVENTION: GALANIN RECEPTOR GALR2 AND
TITLE OF INVENTION: NUCLEOTIDES ENCODING SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:

Tue Jun 1 09:37:58 2004

us-10-063-686-32.fal

Page 8

ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000, 126 E. Lincoln Ave.
CITY: Rahway
STATE: NJ
COUNTRY: USA
ZIP: 07065-0900
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,088A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/033,851
FILING DATE: 27-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 19846
TELECOMMUNICATION INFORMATION:
TELEPHONE: 732-594-1958
TELEFAX: 732-594-4720
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 346 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-993-088A-3

Query Match 4.7%; Score 108.5; DB 3; Length 346;
Best Local Similarity 19.1%; Pred. No. 0.0036;
Matches 81; Conservative 45; Mismatches 124; Indels 173; Gaps 17;

33 WHIFSLVILGLFVCGVLMVLYDYNDLSIELDTERENKCVLGPAIVSTGITAVLLV 92
30 VENFTLVVFGILFAMGVLT-----GNSLVITV 56
93 LIFVLRRIKLTVELFOITNKAISSAPFLF-----QPLMTAILI---FFVWLM 139
57 LARSKRGKRGSTTNLF--ILNLSIADLAVLLFCIPPOATVVALFTVVLGAFICKFIHYFT 115
140 VAVILSLGTAGAAY-----MEGQVEYKPLSGIRYMSYHLIGLWTSFTLA 188
116 VSMVLSIFTLAMSVDRVVAIVHSRRSSSLRVSRNALGVGFMA----- 160
189 COOMTLAGAVTCY---FNRSKNDP-----PDHPILSSLSILFFYHOGTVVKGSPFLIS 238
161 ---LSIAASPAVAYQRLFRHDSNQTFCMEHPNQ-----LHKAYVVCTEVFG 206
239 VVRIPRIIVMOMALKEQHGALSRYLFRCCYCCFWCLDKYLLHLNOMAYTTAINGTD 298
207 -----YLLPILLICF-CYAKVNLHLHK----- 228
299 FCTSAKDAFKILSKNSHFTSINCRGDFIIFLGKVLVVCFTVPGGLMAFNVRAPQVA-- 356
229 -----LKNMSKSEASKKT-----AQTIVLVVVVFG--ISMLEPHVHIMAE 269
357 ---AVPLILVAFAVIAH-----SFLSV-FETVLALFLCFPAVDLETGDS 399
270 FGAFPLTPASFFPRTACLAYSNSVNPITVAILSENFRKAYQVKEKCRVCSNPSGDA 329
400 SEK 402
330 KEK 332
Db

RESULT 13

US-08-993-424B-3
Sequence 3, Application US/08993424B
Patent No. 6337206
GENERAL INFORMATION:
APPLICANT: Tan, Carina
APPLICANT: Kolakowski, Lee F., Jr.
TITLE OF INVENTION: MOUSE GALANIN RECEPTOR GALP2 AND
TITLE OF INVENTION: NUCLEOTIDES ENCODING SAME
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000, 126 E. Lincoln Ave.
CITY: Rahway
STATE: NJ
COUNTRY: USA
ZIP: 07065-0900
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,424B
FILING DATE: 18-DEC-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/033,851
FILING DATE: 27-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Heber, Sheldon O.
REGISTRATION NUMBER: 38,179
REFERENCE/DOCKET NUMBER: 19846NP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 732-594-1958
TELEFAX: 732-594-4720
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 346 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-993-424B-3

Query Match 4.7%; Score 108.5; DB 4; Length 346;
Best Local Similarity 19.1%; Pred. No. 0.0036;
Matches 81; Conservative 45; Mismatches 124; Indels 173; Gaps 17;

33 WHIFSLVILGLFVCGVLMVLYDYNDLSIELDTERENKCVLGPAIVSTGITAVLLV 92
30 VENFTLVVFGILFAMGVLT-----GNSLVITV 56
93 LIFVLRRIKLTVELFOITNKAISSAPFLF-----QPLMTAILI---FFVWLM 139
57 LARSKRGKRGSTTNLF--ILNLSIADLAVLLFCIPPOATVVALFTVVLGAFICKFIHYFT 115
140 VAVILSLGTAGAAY-----MEGQVEYKPLSGIRYMSYHLIGLWTSFTLA 188
116 VSMVLSIFTLAMSVDRVVAIVHSRRSSSLRVSRNALGVGFMA----- 160
189 COOMTLAGAVTCY---FNRSKNDP-----PDHPILSSLSILFFYHOGTVVKGSPFLIS 238
161 ---LSIAASPAVAYQRLFRHDSNQTFCMEHPNQ-----LHKAYVVCTEVFG 206
239 VVRIPRIIVMOMALKEQHGALSRYLFRCCYCCFWCLDKYLLHLNOMAYTTAINGTD 298
207 -----YLLPILLICF-CYAKVNLHLHK----- 228
299 FCTSAKDAFKILSKNSHFTSINCRGDFIIFLGKVLVVCFTVPGGLMAFNVRAPQVA-- 356
229 -----LKNMSKSEASKKT-----AQTIVLVVVVFG--ISMLEPHVHIMAE 269
Db

QY 357 --AVPILLVAFAYVAH-----SFLSY-FETVLDALFLCPAVDLETNDGS 399
DB 270 FGAFPLTPASFFPRTIHAICLAYSNSVNPITIAFLSENFRAXKQVFKRCVNESPHGDA 329
QY 400 SEK 402
DB 330 KEK 332

RESULT 14
US-09-058-333A-5
Sequence 5, Application US/09058333A
Patent No. 6368812
GENERAL INFORMATION:
APPLICANT: Bard, Jonathan A
APPLICANT: Borowsky, Beth
APPLICANT: Smith, Kelli E
TITLE OF INVENTION: DNA ENCODING GALANIN GALR3 RECEPTORS
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/058,333A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52241-E/JPW/XDB
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 278 0400
TELEFAX: 212 391 0525
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 346 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-058-333A-5

Query Match 4.7%; Score 108.5; DB 4; Length 346;
Best Local Similarity 19.1%; Pred. No. 0.0036;
Matches 81; Conservative 45; Mismatches 124; Indels 173; Gaps 17;

QY 33 VHIIFSLVILGLFVCGVLMWLYDYDTNDLSIELDERENMKCVLGPALVSTGITAVALV 92
DB 30 VENFTLVVFGILFAMGVL-----GNSLVITV 56
QY 93 LIFLARKRIKLTVELPQITNKAISSAPFLF-----OPLWTFAILI---FFWVLW 139
DB 57 LARSKRGKRSSTINLF-IINLSIADLAFLFCIPQATVYALPTVVLGAFICKFIHYFT 115
QY 140 VAVLSLGTAGAAOV-----MEGQVEYKPLSGIRVMSYHLIGLWTSFILA 188
DB 116 VSMVLSIFTLAAMSVDRYVAIVHSRRSSLRVRNMLGVGFIMA----- 160
QY 189 CQOMTIGAVVTCY-----FNRKNDP-----PDHPILSLSLFFYHOGTVVKGSPILS 238
DB 161 ---LSIAMSAPVAYYQRLFHRDSNQTFCWEHWPNO-----LHKXAYVVCFTFVG 206
QY 229 VRIIRIIVWMONALKEOQHGLSRYLFRCCYCCFCWCLDKYLIHLNONAYTTTAINGTD 298

DB 207 -----YLLPILICF-CYAKVLNHLNKK----- 228
QY 299 FCTSAKDAFKILSKNSHSPTSGPDIIFLGKVLVCFVFGMLAFNRAFOVW-- 356
DB 229 -----LKNMSKSEASKKKT-----AQTVLVVVVFG--ISMLPHVITHLMAE 269
QY 357 --AVPILLVAFAYVAH-----SFLSY-FETVLDALFLCPAVDLETNDGS 399
DB 270 FGAFPLTPASFFPRTIHAICLAYSNSVNPITIAFLSENFRAXKQVFKRCVNESPHGDA 329
QY 400 SEK 402
DB 330 KEK 332

RESULT 15
US-09-595-549-6
Sequence 6, Application US/09595549
Patent No. 6511827
GENERAL INFORMATION:
APPLICANT: Howard, Andrew D.
APPLICANT: Cascleri, Margaret A.
APPLICANT: Smith, Roy G.
APPLICANT: Sullivan, Kathleen A.
APPLICANT: Tan, Carina
APPLICANT: Van der Ploeg, Leonardus H. T.

TITLE OF INVENTION: GALANIN RECEPTOR GALR3 AND NUCLEOTIDES
TITLE OF INVENTION: ENCODING SAME
FILE REFERENCE: 20148PCA
CURRENT APPLICATION NUMBER: US/09/595,549
CURRENT FILING DATE: 2000-06-16
PRIOR FILING DATE: 1998-12-17
PRIOR APPLICATION NUMBER: US98/26812
PRIOR FILING DATE: 1998-12-17
PRIOR APPLICATION NUMBER: 60/069,725
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 346
TYPE: PRT
ORGANISM: rat
US-09-595-549-6

Query Match 4.7%; Score 108.5; DB 4; Length 346;
Best Local Similarity 19.1%; Pred. No. 0.0036;
Matches 81; Conservative 45; Mismatches 124; Indels 173; Gaps 17;

QY 33 VHIIFSLVILGLFVCGVLMWLYDYDTNDLSIELDERENMKCVLGPALVSTGITAVALV 92
DB 30 VENFTLVVFGILFAMGVL-----GNSLVITV 56
QY 93 LIFLARKRIKLTVELPQITNKAISSAPFLF-----OPLWTFAILI---FFWVLW 139
DB 57 LARSKRGKRSSTINLF-IINLSIADLAFLFCIPQATVYALPTVVLGAFICKFIHYFT 115
QY 140 VAVLSLGTAGAAOV-----MEGQVEYKPLSGIRVMSYHLIGLWTSFILA 188
DB 116 VSMVLSIFTLAAMSVDRYVAIVHSRRSSLRVRNMLGVGFIMA----- 160
QY 189 CQOMTIGAVVTCY-----FNRKNDP-----PDHPILSLSLFFYHOGTVVKGSPILS 238
DB 161 ---LSIAMSAPVAYYQRLFHRDSNQTFCWEHWPNO-----LHKXAYVVCFTFVG 206
QY 229 VRIIRIIVWMONALKEOQHGLSRYLFRCCYCCFCWCLDKYLIHLNONAYTTTAINGTD 298
DB 207 -----YLLPILICF-CYAKVLNHLNKK----- 228
QY 299 FCTSAKDAFKILSKNSHSPTSGPDIIFLGKVLVCFVFGMLAFNRAFOVW-- 356
DB 229 -----LKNMSKSEASKKKT-----AQTVLVVVVFG--ISMLPHVITHLMAE 269

Tue Jun 1 09:37:58 2004

us-10-063-686-32.rai

Page 10

Qy	357	--AVPILLVFPAYLVH-----	SPUSV-FETVLDALFICPAVDLETNDGS	399
Db	270	FGAFPLTPASFFFRITAHCLAYSNSVNPITIYAFLSENFRKAYKQVFKRCVCSNPSFGDA		329
Qy	400	SEK 402		
Db	330	KEK 332		

Search completed: May 27, 2004, 16:38:17
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:32:20 ; Search time 29 seconds
(without alignments)
1476.042 Million cell updates/sec

Title: US-10-063-686-32
Perfect score: 2315
Sequence: 1 MSGRDTTLGLCTIALALSLA.....QDDKSLRNEEGTEQLAVR 445
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	505	21.8	771	2 T16254	hypothetical prote
2	257.5	11.1	523	2 T05692	hypothetical prote
3	179	7.7	488	2 C86385	hypothetical prote
4	177.5	7.7	574	2 T41068	hypothetical prote
5	163	7.0	539	2 S67049	probable membrane
6	126	5.4	372	2 S52054	ubiquinol-cytochro
7	119	5.1	523	2 T11916	NADH2 dehydrogenas
8	118.5	5.1	619	2 T11314	NADH2 dehydrogenas
9	118	5.1	497	1 S53834	NADH2 dehydrogenas
10	117.5	5.0	338	2 S50339	NADH2 dehydrogenas
11	115	5.0	868	2 E89897	conserved hypothet
12	113	4.9	633	2 G95385	Kup2 Potassium upr
13	111	4.8	493	2 S78183	NADH2 dehydrogenas
14	110.5	4.8	694	2 E69143	hypothetical prote
15	110	4.8	495	1 S25942	NADH2 dehydrogenas
16	109	4.7	501	2 B71707	NADH2 dehydrogenas
17	108.5	4.7	608	2 C70111	ADP-ATP carrier pr
18	108	4.7	474	2 B64332	V-type ATPase, sub
19	108	4.7	1783	2 T37258	hypothetical prote
20	108	4.7	1917	2 C88728	probable voltage-d
21	107.5	4.6	447	2 AG0377	protein C48A7.1 fi
22	107.5	4.6	627	2 D90452	probable O-unit fi
23	107.5	4.6	763	2 T48702	hypothetical prote
24	106.5	4.6	501	2 B97790	hypothetical prote
25	106.5	4.6	684	2 T13572	ADP-ATP carrier pr
26	106	4.6	495	2 T12401	NADH2 dehydrogenas
27	106	4.6	497	1 WMB5LM	NADH2 dehydrogenas
28	106	4.6	503	2 D70930	membrane protein l
29	106	4.6	677	2 T50022	hypothetical prote
					sulate transporte

30	105.5	4.6	332	2 B82876	oligopeptide trans
31	105	4.5	782	2 T55925	hypothetical prote
32	104.5	4.5	342	2 T29245	hypothetical prote
33	104.5	4.5	475	2 E73746	MD294 homolog A05-
34	104.5	4.5	492	2 E58931	NADH2 dehydrogenas
35	104	4.5	245	2 B60944	ubiquinol-cytochro
36	104	4.5	445	2 E22845	hypothetical prote
37	104	4.5	544	2 C96943	uncharacterized me
38	103.5	4.5	250	2 C60944	ubiquinol-cytochro
39	103	4.4	446	2 S01187	NADH2 dehydrogenas
40	103	4.4	461	2 S60253	sel-12 protein - C
41	103	4.4	518	2 A53207	probable folate tr
42	103	4.4	602	2 D90557	hypothetical prote
43	103	4.4	1128	2 T04922	hypothetical prote
44	102.5	4.4	306	2 A33960	cytochrome caa3 ox
45	102	4.4	383	2 S55594	G protein-coupled

ALIGNMENTS

RESULT 1

T16254
hypothetical protein F35C8.7 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C/Accession: T16254
R.Wu, X.

submitted to the EMBL Data Library, November 1995
A/Description: The sequence of C. elegans cosmid F35C8.
A/Reference number: Z18486
A/Accession: T16254
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-771 <WUX>
A/Cross-references: EMBL:U40941; NID:g1072184; PID:g1072187; PIDN:AA81709.1; CESP:F35
A/Genes: CESP:F35C8.7
A/Introns: 87/2; 143/2; 238/1; 289/2; 428/2; 629/3; 678/1; 712/1; 736/1

Query Match	21.8%; Score 505; DB 2; Length 771;
Best Local Similarity	28.9%; Pred. No. 2.9e-33;
Matches 131; Conservative	95; Mismatches 180; Indels 48; Gaps 13;
QY	10 LCIALALSLAMTFRRPITTLVHHTISLVILGLVVCGLVW-WLYDYDT--NDLSI 64
DB	318 LIFPAAGILSFITVIMRLGLSLIIMLSILVLVALGAGFSWLMKMTLKTGAIIDYSF 377
QY	65 E--LDTERENMKCVLGFALVSTGITAVLVLIFFLRRIKLTVELFQITNKASAPFL 122
DB	378 HPAADAYEMPTTLVVAIALSVLLFLVILFIRORISLACLISESSKALSSMSTL 437
QY	123 FQPIMTFAIILFFVNLVAVILSIGTAG--AAQVME-GGQVEK-----PLSGIRY 170
DB	438 LFFLPFPLHLHIGVAFALMSIAIWLASSGOEVCRLKETNGQV-VNTSTKCDCTAKVTCQY 496
QY	171 M-----W--SYHLIGLITWSEFLIACQQTINGAVVTCFENSK-NDPPDHILS 217
DB	497 VGIEKESRTITVLOVTLVFAFWISCVTLGLDIALGARASVYAWDKHDVFTFVIR 556
QY	218 SLSLIFFYHOGTVVKGSLISVVRIPRIIVTQMNALKEOQHALSNVLFRCQCCPWCJ 277
DB	557 ALNRAIRYNKISAFGLIIAIVKIIKVLVDHKLGRSQNKAVKMFIM-CLKCCWCL 615
QY	278 DKYLILHNONAYTTAINGDPCISADAKRLSKSSHETSINCFDPIIFIGKLVVC 337
DB	616 EVFFKFLTKNAVIMIALIYGNKPFSSADSFLITRNIVRTVVAKVAGILLFLGKSI-- 673
QY	338 FTVEGGLMAFNVRAPFV-----WAVPLLVAFPAVYVAHGFSLVFETVDALF 386
DB	674 -TLGMLISFYFYSGRVAVGVKVDLYYFVPIVIVISYSYMAIDLFFVYVEMADYTF 732
QY	387 LCFVAVDLETFDGSSEKRYFMDQEFSLFVKSNTL 420

Db 733 ICFLEDSQNDGSLERPFPMSEKLFILGNKNDI 766

RESULT 2

T05692

hypothetical protein F20M13.200 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999

C:Accession: T05692

R:Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.H.

submitted to the Protein Sequence Database, February 1999

A:Reference number: 215420

A:Accession: T05692

A:Molecule type: DNA

A:Residues: 1-523 <BEV>

A:Cross-references: EMBL:AL035540

A:Experimental source: cultivar Columbia; BAC clone F20M13

A:Map position: 4

A:Introns: 264/2; 339/2; 391/1; 437/3

A>Note: F20M13.200

Query Match 11.1%; Score 257.5; DB 2; Length 523;
Best Local Similarity 24.6%; Pred. No. 2,7e-13;
Matches 100; Conservative 75; Mismatches 160; Indels 71; Gaps 16;

QY 6 TILGCIATLALSLAMFFFRITTLVHIFISLVLIGLFCVGLMWLYDYNDISIELDTERENKCVLGA 65
DB 145 TLVVLILISVPCFESVLLKRYTKQIYACPLFLVLPFIF-NYVWVACTLSSGSDA 203
QY 66 LDTRENNKCVLGAIVSTGITALVLLVFLKRIKLTVELFOITNKAISSAPPLFQPL 125
DB 204 LPLAVRIIVLVFVFPII-----GLVWITIVNMRHIDITLQIISVSDALS-----KN 251
QY 126 LMTFAILIFFWLWAVLLSLGTAGAQVME-GGQ--VEYKPLSGIRWMSYHLGLMT 182
DB 252 LKLFVVL-----PLITLGFNKKFVPREIDGQYFCMKEDSVWPYVLAALIMTWS 302
QY 183 SEFIACQOMTIGAVTVCYFNRSKNDP-PDHPILSSLSLFFYHQGTGVKSGSLISVVR 241
DB 303 LAVWEMQVYVIGALIAQWYF--SKEDSIPKKCIRSRNRFQSGFTGICVSGLLICTVR 360
QY 242 IPRIVTMNQMLKQOQHALSRVYFRCCYCCFGLDKYLHMONATTTAINGTDFCT 301
DB 361 VVRAIV---DNARENTQIVWVVL-RCC-----ANALLGEAYCT 396
QY 302 SAKDAFKILSKN--SSHF-----TSINCFGDFIIFLQKVLVVCFTVFGGLMAF---NYNR 351
DB 397 SAKTYELLRRLTSLAVVEVETSTRI-LTGIVFLSAAYAVATWAVLKGVSNGIDISVYV 455
QY 352 AFQVAVFLLVAFVAVYVAHSFSLVSFETVLDALFLCAVLDLTND 397
DB 456 AVLAVWLVLLVILAFVH-----VLDDVIDITVYCAIDRDKGD 493

RESULT 3

C66385

hypothetical protein F20J.7 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001

C:Accession: C66385

R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Comu, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

Nansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 815-820, 2000

A:Accession: C66385
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-488 <STO>
A:Cross-references: GB:AE005172; NID:G10092332; PIDN:AA12743.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 7.7%; Score 179; DB 2; Length 488;
Best Local Similarity 21.1%; Pred. No. 6e-07;
Matches 85; Conservative 62; Mismatches 146; Indels 110; Gaps 18;

QY 21 MFFPRITTLVHIFISLVLIGLFCVGLMWLYDYNDISIELDTERENKCVLGA 80
DB 128 MKSVHILITTYLA-----VSVLCFWCGLFFV-----GGA 157
QY 81 IVSTGTAVALVL-----IFVLRKRIKLTVELFOITNKAISSAPPLFQPLMTFAI 131
DB 158 FAVGSLQFLYVISVIDRLPFTWLVRKALKLWGLPKYVVA-----HAFV 205
QY 132 LIFFWY-LWVAVLLSLGTAGAQVMEGGQVEKPLSGIRWMSYHLGLMTSEFIACQ 190
DB 206 VMLKMSLW---STGAAGVVAASMGDE-----GRWMLVLTSLFMTGAVALCNTV 253
QY 191 QMTIAGAVTVCYFNRSKNDP--PDHPILSSLSLFFYHQGTGVKSGFLISVVRIPRIIV 247
DB 254 HVIYSGWVFVHLFHCQESSSLPESLVDLSRYAVTSGFSIGVSLFTAAIR----- 307
QY 248 MNQMLKQOQHALSRVYFRCCYCCFGLDKYLH-----NQNAVTTAINGTDF 299
DB 308 -----TLRMEIRFRSKICGNELCLC--CVD-FLFLVETLVRFKAVVAVOIAVYKGF 359
QY 300 CTSKADAFKIL-SKNSHFTSINCFGDFIIFLQKVLVVCFTVFGGLMAFVNRARQVMA- 357
DB 360 NKSADAMELPSTGTEALVAAYDCSG-AVLNG-----TIFGLIT--GSCIGINAM 408
QY 358 ----VPLVAFVAVYVAHSFSLVSFETVLDALFLCAVND 392
DB 409 IKYSDRVIVASTAMLMGVLVGLQWVVASVTSIYICFAD 451

RESULT 4

T41068

hypothetical protein SPCC1682.11c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000

C:Accession: T41068

R:Murphy, L.; Harris, D.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, September 1998

A:Reference number: 221968

A:Accession: T41068

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-574 <MUR>

A:Cross-references: EMBL:AL031525; PIDN:CAA2067.1; GSPDB:GN00068; SPDB:SPCC1682.11c

A:Experimental source: strain 972h-; cosmid c1682

C:Genetics:

A:Gene: SPDB:SPCC1682.11c

A:Map position: 3

Query Match 7.7%; Score 177.5; DB 2; Length 574;
Best Local Similarity 21.4%; Pred. No. 9.4e-07;
Matches 91; Conservative 76; Mismatches 159; Indels 99; Gaps 20;

QY 12 IIALAIALAMFFFRITTLVHIFISLVLIGLFCVGLMWLYDYNDISIELDTERE 71
DB 192 IICMMISVIMLFCVLAIPREFLYFLASVPLTWFAFAYVIL-----KASRIHLETSIQ 243
QY 72 NKKCVGFAIVSTG---TAVLVLLFVLRKRI--KLVELFOITNKAISSAP-----FL 121
DB 244 NVNVL-----VPLVAFVAVYVAHSFSLVSFETVLDALFLCAVND 392

QY 122 LFOPLMTFALLIFFWLWVAVLSLGT--AGAAQVMEGGQVEYKPLSGIRYK--SYHL 176
 Db 298 SF--LFSFYLLIFWRLFRRLRGLSTLWGSVWLP-----RSSWVLASFYK 343
 QY 177 IGLWTFEFLAACQMTIAGAVTCEYFNRSKNDPDPHILSLIFNYH-----OGTV 231
 Db 344 LHLMLCTFFHLOQCALISSIVSQWFFYR---DTKSATKTNLVSHFFTHVVSNOYGLCA 400
 QY 232 KGSFLLSVRIPIIIVMKNALKEQHGALSYLFRCYCCGFCMLDKYLLHNONAYTT 291
 Db 401 FSSFLVITVTP--LHFLPTWL-----RHVGRIVY-----YMFSTKTSASYT 440
 QY 292 T-----ATNGDFCTSAKDAFKI-----LSKSSHTSTINCGRDEIIFLGKYLVC 337
 Db 441 SPLTLVASYISVPYMAKALYQIEQNRVGLRRSYTFSK-----ITLLAARSL-- 492
 QY 338 FTVFGGLMAFNY---NRAPQWAVPLLVAFPAVLVASHFLSVFETVLDALFLCPAVDL 393
 Db 493 -AIGGVTSWNYSIHENGVEYGYVG-LTGGFLAWLLIGAIEGSLMIVDALLIGSIIDI 550
 QY 394 ETNDG 398
 Db 551 SSCQG 555

RESULT 5

S67049
 probable membrane protein YOR161C - yeast (Saccharomyces cerevisiae)
 N/Alternate names: hypothetical protein O3568
 C/Species: Saccharomyces cerevisiae
 C/Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 19-Apr-2002
 C/Accession: S67049
 R/Bordone, R.; Camassee, A.; Madania, A.; Martin, R.P.; Poch, O.; Tarasov, I.A.; Wlanc
 submitted to the Protein Sequence Database, July 1996
 A/Reference number: S67032
 A/Accession: S67049
 A/Molecule type: DNA
 A/Residues: 1-539 <BOR>
 A/Cross-references: EMBL:Z75069; NID:g1420398; PID:e252377; PID:g1420399; GSPDB:GN00015;
 A/Experimental source: strain S288C
 C/Genetics:
 A/Gene: MIPS:YOR161C
 A/Cross-references: SGD:S0005687
 A/Map position: 15R
 C/Keywords: transmembrane protein
 F/84-100/Domain: transmembrane #status predicted <TM1>
 F/159-175/Domain: transmembrane #status predicted <TM2>
 F/183-199/Domain: transmembrane #status predicted <TM3>
 F/235-251/Domain: transmembrane #status predicted <TM5>
 F/278-294/Domain: transmembrane #status predicted <TM6>
 F/337-353/Domain: transmembrane #status predicted <TM7>
 F/372-388/Domain: transmembrane #status predicted <TM8>
 F/437-453/Domain: transmembrane #status predicted <TM9>
 F/474-490/Domain: transmembrane #status predicted <TM10>

Query Match 7.0%; Score 163; DB 2; Length 539;
 Best Local Similarity 21.1%; Pred. No. 1.3e-05;
 Matches 94; Conservative 73; Mismatches 160; Indels 118; Gaps 20;
 QY 10 LCTIALSLAMPTEFFITLLVHFISLVIIIGLFFVCGVLMWLVYDYNDSI 64
 Db 136 VCIILAVESVGLTRICRIFPKPI-----YCGAV----- 164
 QY 65 ELDTRENNKCVLGAIV-----STGITAVLVLFV-----LKRRIKLTVELFOI 110
 Db 165 -----INDVASLGRIMWSLRYSAGI--VFLEVTMTACWYGRSRIPLSVAVALKV 216
 QY 111 TKKAISAPFLFQGLMTFALLIFFWLWVAVLSL-----GRAGAQVMEGGQVEYK 163
 Db 217 VVDAAKKCPQIFTFVFGVALVASAGFLFSAYIVATYIKYDPNSNGGCDV--SGGSCSHS 275

QY 164 PLSGIR--YMSYHLIGLWTFEFLAACQMTIAGAVTCEYFNRSKNDPDPHILSL 219
 Db 276 KLIGLVVVPFCGYI-----SEVIRNVICHVSIQVFGSWYMSKSDQGMFPAFGAL 329
 QY 220 SILFPHQGVVKSFLSVRIPIIIVMKNALKEQHGALSYLFRCYCCGFCMLDKYLLHNONAYTT 291
 Db 330 KRAMTYSGSLCPGSLVALIDLRLQILQMTNRHVYTSGGQIA---IQLIFWFDWITIG 386
 QY 279 --KYL--HLNQNAYTTAINGDFCTSAKDAFKILSKN-----SSHFTSI-----NCFG 324
 Db 387 FLKWLAEYFNHVAASFTALYKPYLRAKETWYLRKGMALINDNLINIALGLFSMFA 446
 QY 325 DEIFLGVVYVCT-----VPGIWAFFNNRAFOVW-----AVPLLVAFAYL- 369
 Db 447 SYMTALFTFLYLRFTSPQNSNGAYNGLMAFSVIALQICNATEAIRSGTATFVALG 506
 QY 370 -----VAHSEFLSVFETVLDAL 385
 Db 507 NDPEVFNHSYPRDDEIFRAVPDVL 531

RESULT 6

S52054
 ubiquinol-cytochrome c reductase (EC 1.10.2.2) cytochrome b - Trypanoplasma borreli mi
 C/Species: mitochondrion Trypanoplasma borreli
 C/Date: 14-Jul-1995 #sequence_revision 01-Nov-1996 #text_change 03-Jun-2002
 C/Accession: S52054
 R/Lukes, J.; Arts, G.U.; van den Burg, J.; de Haan, A.; Opperdoes, F.; Sloof, P.; Benn
 EMO J. 13, 5086-5098, 1994
 A/Title: Novel pattern of editing regions in mitochondrial transcripts of the cryptobi
 A/Reference number: S52053; PMID:95045401; PMID:7525275
 A/Accession: S52054
 A/Molecule type: mRNA
 A/Residues: 1-372 <LUK>
 A/Cross-references: EMBL:U11684; NID:g565141; PIDN:AAA73455.1; PID:g565142
 C/Genetics:
 A/Gene: cytb
 A/Genome: mitochondrion
 A/Genetic code: SGC6
 C/Superfamily: cytochrome b, cytochrome b homology; cytochrome b6 homology; plastoquin
 C/Keywords: chromoprotein, electron transfer; heme; iron; metalloprotein; mitochondrio
 F/10-343/Domain: cytochrome b homology <CBH>
 F/125-343/Domain: cytochrome b6 homology <CB6>
 F/183-184/Binding site: heme iron (His) (axial ligands) (low potential) #status predict
 F/197-198/Binding site: heme iron (His) (axial ligands) (high potential) #status predict

Query Match 5.4%; Score 126; DB 2; Length 372;
 Best Local Similarity 20.7%; Pred. No. 0.0094;
 Matches 92; Conservative 64; Mismatches 124; Indels 164; Gaps 24;

QY 10 LCTIALSLAMPTEFFITLLVHFISLVIIIGLFFVCGVLMWLVYDYNDSI 64
 Db 17 LCTILLSGDLFRYVGLGFGNIGWIAQI-----LVGIC--LSMFFRCILPQWTFLL 68
 QY 62 LSTELDTRENNKCVLGAIVSTGI--TAVLVLFVLRKRRIKLTVELFOITKAISAP 119
 Db 69 IHLFFD-----LGFIRSHIIFTSILVFLYLIHIIKIVFLCLIFDS----- 111
 QY 120 FLFQPLMTFALLIFFWLWVAVLSLGT--TAGAAQVMEGGQVEYK-----PLSGIRYMW 172
 Db 112 -----MVMFEGFLIFFILIAF--IGYTLPTSGSYGLTVFSNIALTPIIGI--- 160
 QY 173 SYHLIGLWTFEFLAACQMTIAGAVTCEYFNRSKNDPDPHILSL 219
 Db 161 --YICQWICSEFINFTLLKLSIHIFLPVLLFLIGAHPVLHYF-----L 206
 QY 217 SLSLII-----FPHQGVVKSFLSVRIPIIIVMKNALKEQHGALSYLF 267
 Db 207 SSDGLDRPPFYERFFFLYLRDLFLI--INILCFILYI----- 247
 QY 268 RCCYCCFWCLDKYLLHNONAYTTAINGDF-----FCTSAKDAFKILSKNS 314

Db 248 ----CIYV-----FEVHEBSWIVDTLTKSDKILPEWFFLSFGFLKSPDKF----- 292
 Qy 315 SHFTSINCFDGFIFLGLKLVVCFVTF-----GGMAFNYNRAFOVMAVPLLVAFAVLV 370
 Db 293 -----MGFLPLF-----VLCFALPLFLINCLILITICRSSLIM-MSLSVLFLY-YLC 337
 Qy 371 AHSFLSVFETVLDALFLCFADVLE 394
 Db 338 VCGFLSLY-----VVLCFPLME 355

RESULT 7

T11916
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - Prototheca wickerhamii mitochondrion
 C/Species: mitochondrion Prototheca wickerhamii
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Jun-2002
 C/Accession: T11916
 R/Wolff, G.; Plante, I.; Lang, B.F.; Kueck, U.; Burger, G.
 J. Mol. Biol. 237, 75-86, 1994
 A/Title: Complete sequence of the mitochondrial DNA of the chlorophyte alga Prototheca wickerhamii
 A/Reference number: 217373; MUID:94180393; PMID:8135522
 A/Accession: T11916
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-523 <MOL>
 A/Cross-references: EMBL:U02970; NID:9467843; PID:9467848; PIDN:AAD12635.1
 A/Experimental source: strain HB-8
 C/Genetics:
 A/Genome: mitochondrion
 A/Note: nad4
 C/Superfamily: NADH dehydrogenase (ubiquinone) chain 4
 C/Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 5.1%; Score 119; DB 2; Length 523;

Best Local Similarity 18.1%; Pred. No. 0.049;
 Matches 76; Conservative 63; Mismatches 163; Indels 118; Gaps 15;

Qy 11 CIALATSLAMWTFRRFTTLVHIFSLVIGLLFCVGLMWLYDYNDLSIEDTER 70
 Db 151 CIALVLETMLTVFSVLDLLEFIFESVLIKPIIIGV-WG-----SRRR 196
 Qy 71 ENMKCVGLPAIVSTGITAVALVILFVLRKIKTVELFQTNKAISSAPLLFQPLWTPA 130
 Db 197 KIRAYOFFLYTLFGSVLMLAILIYFOTGTDIEMLYLSDFETQCIIMLAFASFA 256
 Qy 131 IL-----IFW-----VLMVAVLLSISGTAG---AAQWEGQVEXKPSGI 168
 Db 257 VKVMVNVHVLPEAHVBAFTAGSVIILKIKGTGFLRFSPDLPYACIYTPP--- 313
 Qy 169 RYMSYHLIGLWTS-----EFLACQOMTIAGAVTCYFNRSKNDPPHPIISL 219
 Db 314 --IYMSVIAIVYTSCTIRQIDLKIIIVSSVAMNPFVILGLFSQNT----- 359
 Qy 220 SILFFYOGVWKSFLISVVRIRIIVMWQNLKQOQAGLSRYFRCCYCCFWCLDK 279
 Db 360 -----QG--TESILMT-----SHGLVSPALFLCGVLY----- 387
 Qy 280 YLHLNQAAYTTTAINGTDFCTSAKDAFKILSKNSHPTSINCP-----GDPIFLG 331
 Db 388 -----DRHKTRLRLRYSGCGQMPPIFALL--FVEFTMANISLPTSSPGEFLVFIG 437
 Qy 332 KVLVVCFFVGGGLMAFNINRAFOVMAVPLLVAFAVLVAH-SLSVFEVTLDAFLCEA 390
 Db 438 SYQNNISFAVFCATGAVLGAAYALMLCNRLI-----YGVSKPDPINTWSDVNRREFMFA 492

RESULT 8

T11314
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Pedinomonas minor mitochondrion
 C/Species: mitochondrion Pedinomonas minor
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Jun-2002
 C/Accession: T11314
 R/Tume, M.; Lemieux, C.; Burger, G.; Lang, B.F.; Otis, C.; Plante, I.; Gray, M.W.

submitted to the EMBL Data Library, December 1998
 A/Description: The complete mitochondrial DNA sequences of Nephroselmis olivacea and F
 A/Reference number: 217261
 A/Accession: T11314
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-619 <TUN>
 A/Cross-references: EMBL:AF116775; NID:94378766; PID:94378768; PIDN:AAD19666.1
 C/Genetics:
 A/Genome: mitochondrion
 A/Genetic code: SGC3
 A/Note: nad5
 C/Superfamily: NADH dehydrogenase (ubiquinone) chain 5
 C/Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 5.1%; Score 118.5; DB 2; Length 619;

Best Local Similarity 22.0%; Pred. No. 0.064;
 Matches 105; Conservative 73; Mismatches 160; Indels 139; Gaps 24;

Qy 7 ILGLCILALSLAMWTFRRFTTLVHIFIS-----LVTLG 43
 Db 124 ILVAVSENALTLFLGMEIGITSTLLINFWYSKIQSKSAIKAVFLNRIDSPFLNMG 183
 Qy 44 ILFCVGLMWLYDYNDL-----SIEDTERENMK-CVLGPAIVSTGITAVALVILFVLR 98
 Db 184 LTF-----YLFSGDDLFLISLSVFEKQIVKYLIAVLASIAKSAQILHFW-LP 234
 Qy 99 KIKLTVELFQTNKA--ISSAPLLFQPLWTPALILF-----WLMVAVLLS--GTAG 150
 Db 235 DAIEAPTPVSSLHATVAGAVYLIR--LSFLSDFDSSNFIIVIGITSLFAGIG 292
 Qy 151 AAQWEGQVEXKPSGISRYMSYHLIGL-----WTSEFLACQOMTIAGAVTCY 202
 Db 293 FNOPTKRIIVSTCSQIGM--FYALGISLSDPSYHLFVHGFKMLFLAAGFIITI 350
 Qy 203 FNRSKNDPPDP-----ILSSSLT--FFYHGTGVKSGFLISVVRIRIIVM 248
 Db 351 FN--EQDIRKYSGLNPLNSFLSAFVLSNLSLGIFFAGFYSKELLISGI----- 400
 Qy 249 WQNLKQOQAGLSRYFRCCYCCFWCLDKILHLNQAAYTTTAINGTDFCTSAKDAFK 308
 Db 401 YTN-----FWSL-----LSVLAFTTCVY-----IKSILL 427
 Qy 309 IL--SKNSHPTSINCFD--FIIFLGLVAVVCFVTFG-----GLMAFNYNRAFOVMA 357
 Db 428 VLSGSPNNMSPSITTYTNILVILISLVAVLNLFPEPMITEQIKMDIYVNS----- 482
 Qy 358 VELLVAFPAYVANSFISVFEVTLDAFLCPADVLEINDDSSSEKPYMDDEFLSFV 414
 Db 483 -----VPRS--IKSNFLFEHYVYVNIILFVLSVFLISLWNSYIFPFSKLSFI 530

RESULT 9

S53834
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - Acanthamoeba castellanii mitoc
 C/Species: mitochondrion Acanthamoeba castellanii
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002
 C/Accession: S53834
 R/Burger, G.; Plante, I.; Loneragan, K.M.; Gray, M.W.
 J. Mol. Biol. 245, 522-537, 1995
 A/Title: The mitochondrial DNA of the amoeboid protozoon, Acanthamoeba castellanii: cc
 A/Reference number: S53825; MUID:95147275; PMID:7844823
 A/Accession: S53834
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-497 <BNA>
 A/Cross-references: GB:U12386; NID:9562028; PIDN:AAD11826.1; PID:9562038
 A/Experimental source: strain Neff; ATCC 30010
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
 C/Genetics:
 A/Genome: mitochondrion
 A/Genetic code: SGC6
 C/Superfamily: NADH dehydrogenase (ubiquinone) chain 4

C:Keywords: membrane-associated complex; mitochondrion; NAD, oxidative phosphorylation;

Query Match 5.1%; Score 118; DB 1; Length 497;
Best Local Similarity 19.5%; Pred. No. 0.056;
Matches 90; Conservative 65; Mismatches 143; Indels 164; Gaps 22;

```

QY 11 CIALALSLAMMFFRITTLVHIFSLVILG-----LFFVCGV 50
DB 140 CVALD-----VEFYIFESILIPMFILIGWGRORKHAVYOLFPYLLGSLMLLGI 193
QY 51 LMMVYDYNDLSIELDERENKCVLGAIVSTGTAVALVLIIVLARKIKLVELPQI 110
DB 194 LVYISHQTDIVLVNTN-----FSFYRLILMASPF-- 226
QY 111 TNKAISAPPLFQPLMTFAILIFW-----VLWVAVLISGTAGAAQWBG 157
DB 227 -----FACVAVPLFPEFHI--MLPEAVPEPTVGSVILAGVLKLGTYGL----- 269
QY 158 GQVEYKPLSGIRYV-----WSYHLGLWTSSEFILAQQMTAGAVTCYFRSKNDP 211
DB 270 -----LRVVIPLFCDAITFPLVYT-----LCMLGITVCCSTTRQVDLK 310
QY 212 DHDILSLILFFYHOGTVKGSFLISVARIPIIWMYQNALKEQOHALSRYLEFCY 271
DB 311 KVIAYASVS-----HMSFVILGLFTSNIGIGSVFLML-----SHGIVSGLFPCIG 358
QY 272 CCWCMDKY--LHINQNAVTTAINGDPCTSADAPEKILSKNSHFTSINCQDPII 328
DB 359 CYV--DRYKTRILRYSGLVSTMP--FSLC-----LPLILSNISFPGTSFICEFVI 408
QY 329 FLG-----KVLVCFVTEVGGIMAFVYRAFOVAVPPLVAVFAVYVAHSPL-----S 376
DB 409 LGLFENNHPALIAFSTI-----LNAVSIWLYNRIT--FRLVAVNYILRPSDS 458
QY 377 VERTVDALFLCPAVDLETNDG-----SEKPYMDQEFLSF 413
DB 455 KKEFVVGTF-CPTITLFGIKGSYIILIEAFIV--YLSF 496

```

RESULT 10

NRH32 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - Katharina tunicata mitochondrion

C:Species: mitochondrion Katharina tunicata

C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 03-Jun-2002

C:Accession: S50339

R:Boore, J.L.; Brown, W.M.

Genetics 138, 423-443, 1994

A>Title: Complete DNA sequence of the mitochondrial genome of the black chiton, Katharin

A:Reference number: S50327; MUID:95129806; PMID:7828825

A:Accession: S50339

A:Molecule type: DNA

A:Residues: 1-338 <BOO>

A:Cross-references: EMBL:U09810; NID:g557273; PIDN:AA48374.1; PID:g335326

A>Note: the authors translated the initiation codon GTG for residue 1 as Val

C:Genetics:

A:Gene: ND2

A:Genome: mitochondrion

A:Genetic code: SGC4

A:Start codon: GTG

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 2

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 5.1%; Score 117.5; DB 2; Length 338;

Best Local Similarity 21.6%; Pred. No. 0.042;

Matches 85; Conservative 52; Mismatches 120; Indels 137; Gaps 20;

```

QY 22 MPTFRITTLVHIFSLVILGLFVCGVLMVLYDYNDLSIELDERENKCVLGA 81
DB 1 MNPFP-----VGLIFILIFFTLPSLSIHW-----FGVWGLLEIN-----LMGF-- 41
QY 82 VSTGTAVALVLIIVLARKIKLT--VELFOITNKAISAPPL--LFPPLMTFAI----- 131
DB 42 -----IPVWVOKSTSEETESGVKYLIV--QAVGSAFLFLPGLMAMNMFCCWELNF 89

```

```

QY 132 -----LFFWVLMVAVLLSLGTAGAAQWEGQVEYKPLSGIRYVMSYHLI-GLIWT 182
DB 90 FSGFSGXSLIIF-----GLMKLGA-----P-----HWVSVVAGLSWM 126
QY 183 SEFIIACQOMTAGAVTCYFRSKNDPDPHILSLISILFFYHOGTVKGSFLISVRI 242
DB 127 SNFLILFVQ-KIAPLFVWCYIN-----LSSFLILIVELSSIDFGVGQVQTSV 175
QY 243 PRII-----VMYQNALKEQOHALSRYLEFCYCCFPCDLYLHINQNAVTTAINGD 298
DB 176 RALIVYSSILHGMWMLKASAGWSITFFYFFYCFPIIGFAYLMLGDE----- 223
QY 239 FCTSAKDAFKILSKNSHFTSI--NCFG-DPIIFLGKVLVVCPTVFGGLMAFVYRAFO 354
DB 224 -----SFVMSCFSSVYVWNGYSRFLVFM-----LLSIGL----- 254
QY 355 VNAVPLIVAFPAIVVAHSFISVEFYLDAFLC 368
DB 255 -----PPLIGFQKMLVAVSLISLGNLVLSITLVC 284

```

RESULT 11

B89897

conserved hypothetical protein SA1085 [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: B89897

R:Kurdoz, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Og

ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: B89897

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-868 <KUR>

A:Cross-references: GB:BA000018; PID:g13701042; PIDN:BA842337.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: SA1085

Query Match 5.0%; Score 115; DB 2; Length 868;

Best Local Similarity 18.6%; Pred. No. 0.17;

Matches 73; Conservative 63; Mismatches 140; Indels 116; Gaps 16;

```

QY 8 LGLCIALLALSLAMMFFRITTLVHIFSLVILGLFVCGVLMVLYDYNDLSIELD 67
DB 197 IGRIVAILTLFNSFYFS-----YQALIIIG-----CYLVKILFYYKXDIV----- 239
QY 68 TERENKCVLGAIVS-----TGITVVL-----LVLIIVLARKIKLVEL 107
DB 240 SRQKFLCVASATVLSLSSVFGLFGLISAFLENDRKONPNVDIPILTPADYHYFFESDG 299
QY 108 FOITNKAISAPPLFQPLMTFAILIFFWLVAVVLLSGTAGAAQWEGQVEYKPLSG 167
DB 300 FYIILISILITVALLSPGLYFYFRRLPAIVTWLIFGSL-----SQVPSDA--FNGFSF 351
QY 168 IRYKSYHILGLIWTSEFIIACQOMTAG--AVTCYFRSKNDP----- 210
DB 352 PERRVWV-----LIALSSALCGFIQHLSTLNMKYLIIRTPVSIYALVILL 400
QY 211 -PDHPLISSILFFYHOGTVKGS-----FLISVVRIRPIIWM-----Y 249
DB 401 SPFHPLALYGIILIMVLAIVLKFSLMRKGLTVALIVLIVMIOQLVIDNNKZMAIKPE 460
QY 250 MKNALKEQOHALSRYLEFCYCCFPCDLYLHINQNAVTTAINGDPCTSADAPEK 309
DB 461 QQSILSTLKQHDYHSN-----VQILKIKINQNA--TGFPRIDYMSD-----YAL 503
QY 310 LSKNSHFTSINCF-----GDPIIFLGKVLV 336

```

Db 504 NSPFIHYNGISLYSIFNGDILKYDXTLQI 535

RESULT 12

Kup2 Potassium uptake protein [imported] - Sinorhizobium meliloti (strain 1021) magaplas
C/Species: Sinorhizobium meliloti
C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C/Accession: G95385

R/Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows,
Kalmann, S.; Keating, D.H.; Kalm, C.; Beck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A/Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
A/Reference number: A95262; MUID:21396509; PMID:11481432
A/Accession: G95385

A/Status: preliminary
A/Molecule type: DNA

A/Residues: 1-633 <KIR>
A/Cross-References: GB:AE006469; PIDN:AAK65649.1; PID:G14524136; GSPDB:GN00165
A/Experimental source: strain 1021, megaplaemid psymA
R/Gallbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Barloy-Hubler,
Pela, D.; Chahn, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001

A/Authors: Kahn, D.; Kahn, M.L.; Kalmann, S.; Keating, D.H.; Kise, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A/Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A/Reference number: A96039; MUID:21368234; PMID:11474104
A/Contents: annotation

A/Genetics:
A:Genome: kupd2
A:Genome: plasmid

Query Match 4.9%; Score 113; DB 2; Length 633;
Best Local Similarity 22.0%; Pred. No. 0.18;
Matches 49; Conservative 39; Mismatches 67; Indels 68; Gaps 9;

QY 3 GRDTIIGLIALALSLAMFFRTITTL-----VAIFSLVILG-----LHVV 47
DB 54 GREVEIGLVSIVL-WITTAIVTITKYVLFILRADNDEGGTLSTLALLKKKTKYPTLMFF 112
QY 48 CGVLMWLY-----DYTNDSIEIDTERENKCVLGFALVSTG-----ITAVLVLFV 96
DB 113 AGVLGALFLIGDAMITPPLSV-----LSNVEGLVAVPALADVVLISVITILLRA 164
QY 97 LRRIKLTVEFLQITNKATISAPFLFQPLMTFAIIFFWLVAVVLSLGTGAQAQVME 156
DB 165 VQGR-----GTGAVSVFFQPIITLWPFLLMAAAGVNHID 198
QY 157 GGGV--EYKPLSGIRVMSYHLGLTWSEFILAQQOMTLAG 197
DB 199 DLALISAFNPLNIGFLMNAGLIGFT-----VLGAFLVTGTA 236

RESULT 13

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - Recclinomonas americana (ATCC 503
C/Species: mitochondrion Recclinomonas americana
A/Variety: ATCC 50394
C/Date: 26-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 03-Jun-2002
C/Accession: S78183

R/Lang, B.F.; Burger, G.; O'Kelly, C.J.; Cedergren, R.; Golding, G.B.; Lemieux, C.; Sank
Nature 387, 493-497, 1997
A/Title: An ancestral mitochondrial DNA resembling a eubacterial genome in miniature.
A/Reference number: S78127; MUID:97311393; PMID:9168110
A/Accession: S78183

A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-493 <LAN>
A/Cross-References: EMBL:AF007261; NID:G2258325; PIDN:AAD11916.1; PID:G2258382
A/Experimental source: ATCC 50394

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1997
C/Genetics:

A/Genome: nad4

A/Species: mitochondrion
C/Species: NADH dehydrogenase (ubiquinone) chain 4
C/Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation

Query Match 4.8%; Score 111; DB 2; Length 493;
Best Local Similarity 21.8%; Pred. No. 0.21;
Matches 85; Conservative 53; Mismatches 128; Indels 124; Gaps 21;

QY 10 LCTALALSLAMFFRTITTLVHIFISVILGLFVCGVLMVLYDYDNDLSIEDTE 69
DB 120 ICFLLDALLIMIFC--VLDELVFYIFPESVLIEMFIVIGV-WG-----SRE 163
QY 70 RENNKCVPALVSGITAVLVILFV-----RRIKLTVEFLQITNKATISAPFL 122
DB 164 RYVRAVMLFLYTFPG-SLMMLALVWTFPDAGTTDQVLTTFPSORQLWLAFFIS 222
QY 123 PQ--PLWTFALILFFW-----VLWVAVLLSCTAG-----AAQWEGQVVE 162
DB 223 FAIRKIPWPFHV---MLPEAHEAPTAGSVLGLVGLKGGYGLRFSIMPEEATVVF 278
QY 163 KPLSGIRVMSYHLGLTWSEFILAQQOMTLAGAVVTCYRNKNDPDPHILSSIL 222
DB 279 TPL-----VYMSITAIYTS--LTLRQIDLKRIIA-----YSSVA-- 313
QY 223 FFYHGGTVVKSFLISVVRIPRIITWYQNALKEQHGALSRYLFRCCYCCFCLDKY-- 280
DB 314 --HNFTIGMFLMNOGLESLIML-----SHGIYSSALFLICGVLY--DNKT 360
QY 281 -LHINQNAVYTTIANGTDFC--TSADAFKILSKNSHFTSINCPGDFIIFGKVLVVC 337
DB 361 RLKXYSGVQVQTMPIFALFPLFTMA-----NISLPTGSFV-----GEFVLIG-- 405
QY 338 FTVFGGLMAFVYNRAFOVWAVPLLVAFPA 367
DB 406 -----AFNSNTT-----VAFPA 417

RESULT 14

E69143
Hypothetical protein MTH339 - Methanobacterium thermoautotrophicum (strain Delta H)
C/Species: Methanobacterium thermoautotrophicum
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C/Accession: E69143

R/Smith, D.R.; Doucet, L.A.; Delughery, C.; Lee, H.; Dubois, J.; Aldredge, T.
X.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani,
J.; Bacteriol. 179, 7135-7155, 1997

A/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fur
A/Reference number: A69000; MUID:98037514; PMID:93711463
A/Accession: E69143

A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-694 <MTH>
A/Cross-References: GB:AE000818; GB:AE000666; NID:G2621384; PIDN:ABB84845.1; PID:G26213

A/Experimental source: strain Delta H
A/Genome: MTH339

Query Match 4.8%; Score 110.5; DB 2; Length 694;
Best Local Similarity 20.5%; Pred. No. 0.22;
Matches 88; Conservative 73; Mismatches 164; Indels 105; Gaps 20;

QY 12 LIALALSLAMFFRTITTLVHIFISLV-----LGLFVCGVLMVLY----- 56
DB 58 VTAALSLIFLM---MWTGLINSFPLINRPSLAPISAINLLIT--ALNVIYHURGE 111
QY 57 -----DYTNDSIEIDTERENKCVLGFALV--TGITAVLVILFVLRRIKLTVEFLQ 109
DB 112 EPLNWKASGNLSTSLSPFLVLTGVSYLMNKSVNTVLLVLF-----IVYITLLE 166
QY 110 ITNKATISAPFLFQPLMTFAIIFFWLVAVVLSLGTGAQAQVMEGGVEKPLSGIR 169

```

Db      167 VFKKKIGPA-----TYPVAF-----SISLSL-----LIMNGLEPSNY--LIGRD 203
Qy      170 YMSGVHILGL-----IMTSEFLAACQMTIAGAVTTCYFNKSKNDPPHPLISL 219
Db      204 IHMEFYLRKALMAHMDMVGSNAYNACLSTVLEVIYKULN-----VPEY 252
Qy      220 SILFFYHOGTVKSSFLISVVRIPRIIVMTQNALKEQHGALSRYL----FECYCCFW 275
Db      253 YIFKYY-----GFIGALMPLP--VLLISERILKSDYGFYATLLFIQFSGFIYLGW 303
Qy      276 CLDKYLLHNGNAVTTTAINGTDFCTSAKDAFKILSKNS--SHPSTINCGDPIIFLGKY 333
Db      304 C--RQVALLFPFAAAMVLTGDMRSHKLLFVFMVGYLSHITTAVF---FELVAL 357
Qy      334 LVGCTVFGGLMAFNRAF--QVAVPPLLVAFAVLA----HSFISVPEVLDALF 386
Db      358 IPIVVRVKRLKVPDSDSGFFNASLAVLFVVVFAWYAQTGAPFKSAVSFFTKLRMS 417
Qy      387 LCPAVDLETN 396
Db      418 EFSHDMNN 427

```

RESULT 15

S25942

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - liverwort (Marchantia polymorpha)

N.Alternate names: NADH-ubiquinone oxidoreductase chain 4

C.Species: mitochondrion Marchantia polymorpha

C.Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002

C.Accession: S25942; S31172

R.Oda, K.; Yamato, K.; Ohta, E.; Nakamura, Y.; Takemura, M.; Nozato, N.; Akashi, K.; Kan

J. Mol. Biol. 223, 1-7, 1992

A.Title: Gene organization deduced from the complete sequence of liverwort Marchantia po

A.Reference number: S25941; MUID:92114051; PMID:1731062

A.Accession: S25942

A.Status: nucleic acid sequence not shown; translation not shown

A.Molecule type: DNA

A.Residues: 1-495 <ODA>

A.Cross-references: EMBL:M68929; NID:9786182; PIDN:AC09398.1; PID:9786185

A.Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992.

R.Noizato, N.; Oda, K.; Yamato, K.; Ohta, E.; Takemura, M.; Akashi, K.; Fukuzawa, H.; Oh

Mol. Gen. Genet. 237, 343-350, 1993

A.Title: Co-transcriptional expression of mitochondrial genes for subunits of NADH dehyd

A.Reference number: S31171; MUID:93247547; PMID:8483448

A.Accession: S31172

A.Status: nucleic acid sequence not shown

A.Molecule type: DNA

A.Residues: 1-495 <NO2>

A.Cross-references: EMBL:M68929; NID:9786182; PIDN:AC09398.1; PID:9786185

C.Genetics:

A.Gene: nad4

A.Genome: mitochondrion

A.Introns: 183/2

C.Superfamily: NADH dehydrogenase (ubiquinone) chain 4

C.Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 4.8%; Score 110; DB 1; Length 495;

Best Local Similarity 18.9%; Pred. No. 0.25;

Matches 78; Conservative 73; Mismatches 152; Indels 110; Gaps 16;

Qy 12 ILAALSLAMMTFRPTTLIVHIFISLVILG-----LTFVGVV 51

Db 130 LIAVFCSLDLIFVFFESVLIIPMFIIIGWGSROKIKAYOFPLYTLGSLFMDLAIL 189

Qy 52 WMLVYDYNDISILDETERENMKCVLGAFAIVSTGITAVLVLLFVLRKRIKLVESLQIT 111

Db 190 FIFQGTGTDQILITE-----FSERRQILMLIAFFAS 224

Qy 112 NKAISAPFLLFQPLMTFAIIFFWVLMVAVLISLGTAG---AAQVMEGGQVEYKPLSG 167

Db 225 SVKVPWVFVHILWLSAHEAVEAFAGSVILAILKLTGYGFLRPSIPMFPEATLYFTP--- 281

Qy 168 IRYMWSYHILGLIWTSEFIILACQMTIAGAVTTCYFNKSKNDPPHPLISLSLIFPYHQ 227

```

Db      282 ---FVTLSTVIAIYTS--LFTIRQIDKRIA-----YSSVA-----HM 316
Qy      228 GTVYKGSFLISVVRIPRIIVMTQNALKEQHGALSRYLFRCCYCCFWLGDY---LHL 284
Db      317 NFVTIGKFSINTQIGESILML-----SHGLVSSALFLCVGALY---DNRKTRIVKY 366
Qy      285 NONAVTTTAINGTDFCTSAKDAFKILSKNSHPTISNCGDPIIFLG---KVLVVCFTV 340
Db      367 YGGLVSTMPISFTIF-----LFTLANMSLPOTS-SFIGEFLLIVGAFQRSIVATLAA 419
Qy      341 FGLMANNRARAQVAVPPLLVAFA--YVAHSLSLSEFTVLDALFECFAV 391
Db      420 LGMIL---GAAYSLMDYNRVFNGFNPNFLKFSDLNRREV---IFLPFTV 465

```

Search completed: May 27, 2004, 16:37:42

Job time : 31 secs